WATER CUT ESTIMATION IN HIGH GOR WELLS

BY

MOHAMMAD SAEED AL-KADEM

A Thesis Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

PETROLEUM ENGINEERING

DECEMBER 2018

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Mohammad Saeed Al-Kadem** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN PETROLEUM ENGINEERING.**

Dr. Dhafer A. Al-Shehri
(Advisor)

Dr. Dhafer A. Al-Shehri
Department Chairman

Dr. Mohamed A. Mahmoud
(Member)

Dr. Salam A. Zummo
Dean of Graduate Studies

Dr. Rahul N. Gajbhiye
(Member)

17/12/18

Date

Dedication

To my parents, precious wife, son, and upcoming daughter

ACKNOWLEDGMENTS

I want to thank ALMIGHTY ALLAH for his unlimited support. I would also like to thank my parents who looked after and helped to educate me very well since my birth. I want to extend appreciation and again dedication to my precious wife and precious son for their patience. I could not have reached this far without them motivating me and helping me to get rid of all obstacles on the way to my graduation. I would also like to send special thanks to the thesis committee advisor, Dr. Dhafer Al-Shehri, and thesis committee members, Dr. Mohamed Mahmoud and Dr. Rahul Gajbhiye, for their guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
CAPEX	Capital Expenditure
CC	Correlation Coefficient
EOR	Enhanced Oil Recovery
ESP	Electric Submersible Pump
GOR	Gas-Oil Ratio
GVF	Gas Volume Fraction
ICV	Inflow Control Valve
MPC	Multiphase Flow Correlations
MPFM	Multiphase Flow Meter
NN	Neural Network
RF	Random Forest
WC	Water Cut

ABSTRACT

Full Name : Mohammad Saeed Al-Kadem

Thesis Title : Water Cut Estimation in High GOR Wells

Major Field : Petroleum Engineering

Date of Degree : December 2018

Water-cut measurement is important for reservoir management and recovery maximization. It is also considered as a key parameter to monitor well performance. Nevertheless, the water-cut measurement is severely affected when a large volume of gas comes out of solution in high Gas-Oil Ratio (GOR) reservoirs. Inaccurate water-cut measurement could be due to metering technologies failure when Gas Volume Fraction (GVF) exceeds 90%. Therefore, the objective of this research is to develop a model (s) to estimate water cut better in wells with GOR values higher than 2,000 scf/STB.

One of the conventional methods used in measuring water cut is the multiphase flow meter (MPFM). This equipment gives accurate and reasonable values at conditions, where there are no flow assurance issues at the surface, such as free gas, slug, or emulsions. MPFMs are costly and prone to failures. Nevertheless, there are few metering technologies applicable to a high GOR environment.

In this research, two models are developed based on nonlinear multiple regression and an artificial neural network (ANN). The developed models utilized various inputs that are measured on a real-time basis taken from surface and subsurface sections of the well to better estimate water cut in high GOR wells.

Multiphase flow correlations and models are available in the literature to provide an estimation of water cut with limitations in the GOR range. Some multiphase flow correlations were selected based on GOR applicability and compared to actual water cut measurement with an average absolute relative percentage error of 25%. A new empirical correlation was developed using nonlinear multiple regression. Results showed that the correlation is more reliable at low GOR with an average absolute relative percentage error of 8.32%. Nevertheless, another model using ANN was developed covering a wide range of GOR. It demonstrated more reliable results with an average absolute relative percentage error of 8.13% when comparing their values to the actual water cut from test separators. This improvement in water-cut estimation was due to the use of a wide range of surface and subsurface parameters, provided that they are obtained on a real-time basis.

The estimation of an accurate water cut is imperative for better management of reservoirs in case of abnormal conditions such as high GOR. Practicable and cost-driven models were developed to estimate water cut in the absence of reliable metering technology or robust multiphase flow correlations for high GOR application.

ملخص الرسالة

الاسم الكامل : محمد سعيد آل كاظم

عنوان الرسالة : تقدير نسبة الماء في الآبار التي فيها نسبة الغاز عالية

التخصص : هندسة بترول

تاريخ الدرجة العلمية : ربيع الثاني 1440 هجري

إن قياس نسبة إنتاج الماء المصاحب لإنتاج الآبار النفطية ضروريٌ ، خاصة عندما يقترن ذلك بإنتاج الغاز ، مما يجعل قياس نسبة كل الموائع (الماء ، النفط ، الغاز والغاز السائل) أمرًا صعبًا. تتأثر دقة قياس نسبة المياه بشدة عندما تكون مصحوبة بنسب كبيرة من الغاز المصاحب. ولذلك ، فإن الهدف من هذا البحث هو تطوير طرق لتقدير نسبة المياه في آبار نفط مصاحبة لنسب غاز عالية.

واحدة من الطرق التقليدية في قياس نسبة الماء هي مقياس التدفق للهيئات الفيزيائية. توفر هذه التقنية قيمًا دقيقة ومعقولة في الظروف النموذجية ، عندما لا يكون إنتاج نسبة الغاز عاليًا. ومع ذلك ، فإن أجهزة قياس التدفق للهيئات الفيزيائية غالية ، مكلفة وتستلزم صيانةً بشكل دوري . ومع ذلك ، هناك عدد قليل من تقنيات القياس لقياس نسبة الماء في الآبار التي فيها نسبة عالية من الغاز.

في هذا البحث ، تم تطوير نموذجين بإستخدام الانحدار المتعدد غير الخطي والشبكة العصبية الاصطناعية. استخدمت النماذج المتطورة مدخلات مختلفة تقاس في الوقت الحقيقي مأخوذة من أقسام السطح والجزء تحت السطحي من البئر ، لتقدير أفضل لنسبة الماء في آبار التي فيها نسبة الغاز العالية.

تتوافر العلاقات المترابطة ونماذج التدفق للهيئات الفيزيائية في مصادر البحث المتوفرة بتقديم تقدير لنسبة المياه بحدود معينة. قد تم اختيار بعض من هذه العلاقات المترابطة ومقارنتها مع قياس نسبة المياه الفعلية مع متوسط الخطأ بنسبة مئوية مطلقة تبلغ 25 ٪. لذلك ، تم تطوير علاقة جديدة باستخدام الانحدار المتعدد غير الخطي. وأظهرت النتائج أن الارتباط أكثر موثوقية عندما تكون نسبة الغاز قليلة مع متوسط نسبة الخطأ تبلغ 8.32٪. بناءًا على ذلك، تم تطوير نموذج آخر باستخدام الشبكة العصبية الاصطناعية يغطي مجموعة واسعة من نسبة الغاز. وقد أظهرت نتائج أكثر دقة مع وجود متوسط الخطأ بنسبة مئوية مطلقة تبلغ 8.13٪ عند مقارنة قيمها بنسب المياه

الفعلية المقاسة من أجهزة الحقول الذكية. ويعزى هذا التحسن في تقدير نسبة المياه إلى استخدام مجموعة واسعة من المتغيرات السطحية والجوفية.

يعد تقدير نسبة للمياه بدقة أمرًا ضروريًا لتحسين إدارة الحقول والمكامن في حالة الظروف غير الطبيعية مثل نسبة الغاز والنفط المرتفعة. تم تطوير نماذج عملية وغير مكلفة لتقدير نسبة المياه ، في غياب تقنية قياس موثوق بها أو علاقات موثوقة لتدفق الهيئات الفيزيائية للآبار المصحوبة بنسب غاز عالية.

CHAPTER 1 INTRODUCTION

1.1 Background

For any technical research, there has to be one of three areas to consider; solving a problem, improving a solution or discovering a new challenge. The objective of the research should capture one or more of the mentioned areas. The oil industry business is dynamic; new issues are discovered and solutions need to be addressed and further improved. For instance, production of oil from tight sands might be considered a problem. Oil companies and services providers are working collaboratively to address this problem and bring the most innovative technologies and methods to solve it. Another area of research includes a new empirical correlation developed to estimate gas flow rate with an error of 10%. A new solution used Artificial Intelligence (AI) techniques, reducing the error to 5%, which is called a solution improvement. In this research, a solution is to be addressed using available numerical techniques such as, nonlinear regression modeling or AI methods to accurately estimate water-cut at a high gas-oil ratio (GOR), which profoundly affects rates and water-cut measurement.

Water-cut is defined as the ratio of water flow rate to the liquid flow rate in percentage (Economides, 1996).

$$Water - cut\% = \frac{Q_w}{Q_o + Q_w} x100 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1.1)$$

Field or well rate measurement are done by using one of the following metering technologies:

- Single-phase flow meters such as the Weatherford Red Eye water-cut meter
- Permanent or portable test separators

- Multiphase flow meters (MPFMs)

The standard classification of reservoir fluids below is based on the gas-oil ratio in **Table 1.1** (McCain, 1990):

Table 1.1 Classification of reservoirs based on the gas-oil ratio in scf/STB

Black Oil	Volatile Oil	Gas Condensate
< 2,000	2,000 to 3,300	3,300 to 50,000

Gas-Oil Ratio (GOR) is defined as the gas flow rate over the oil flow rate.

$$GOR = \frac{Q_g}{Q_o} \quad , \qquad scf/STB \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

This is for the type of reservoir that is considered as being volatile oil or gas condensate. Therefore, wells with a GOR value of more than 2,000 scf/STB are deemed to be **high GOR wells**.

1.2 Importance of Water-cut Estimation

I. Impact of High Water Cut and High GOR on Well Performance

Water-cut measurement becomes essential in fields with high water production. Oil companies tend to produce oil and maximize recovery avoiding the excessive production of water where this necessitates monitoring water cut at the early life of the well to prevent a scenario of producing additional water provided that accurate measurement of water cut is attained.

Accurate measurement will enable decision makers to advise to shut-in or sidetrack wells with high water cut to maintain the oil production of the field (Ehtesham, 2011). Also, reservoir simulation will require accurate water cut as an input way to ensure the precision of the simulation model for engineers to forecast field production and advise any additional required action that would sustain or enhance oil recovery. Water cut profoundly affects well inflow performance especially when the oil-water contact rises to the oil zone. **Figures 1.1 and 1.2** demonstrate examples of a low GOR well with the effect of water cut. **Figure 1.1** shows an example of a well with low GOR. Although the water cut increased by 10% during a period of two years, this did not have a substantial impact on oil production. In contrary, in **Figure 1.2**, water-cut is increasing with time causing an appreciable decline rate in oil production in low GOR wells.

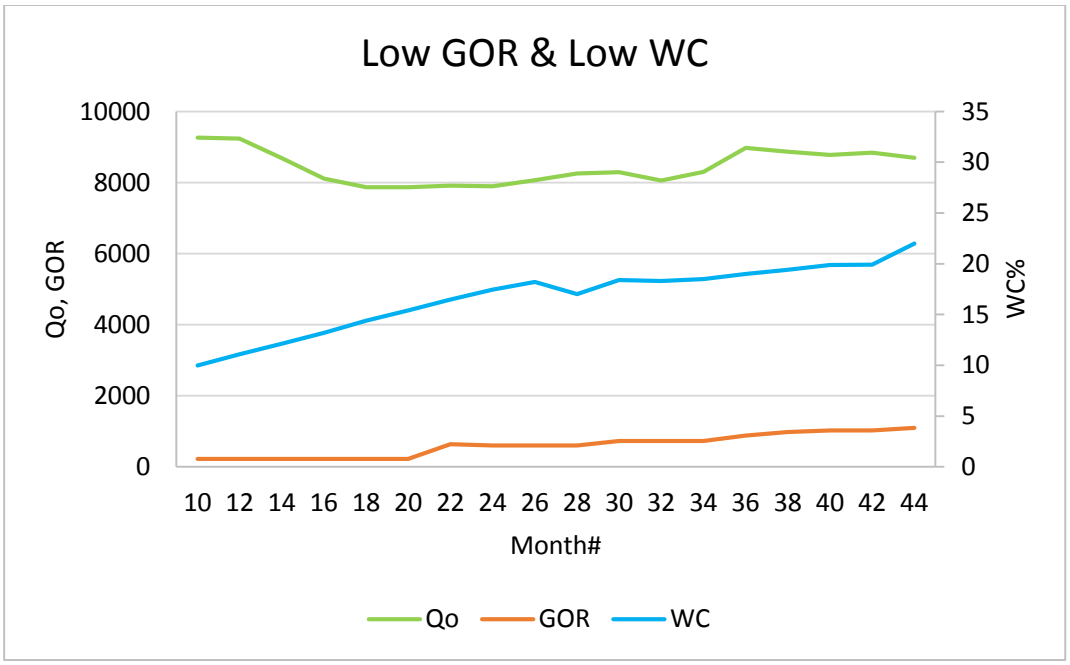


Figure 1.1 Gas oil ratio and water cut impact on well performance (low GOR and low WC)

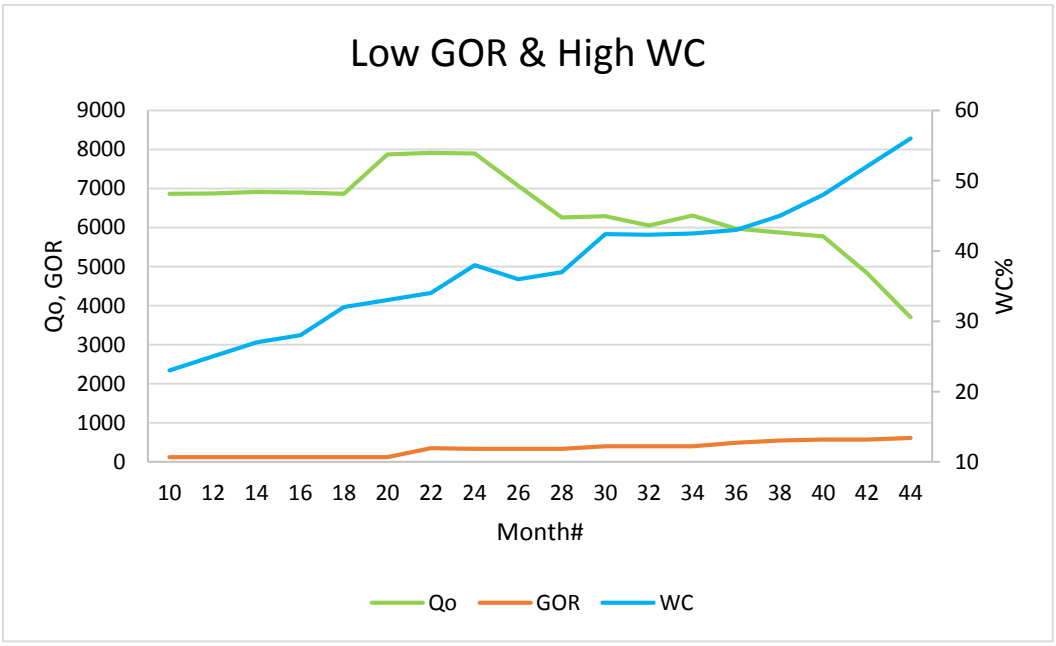


Figure 1.2 Gas oil ratio and water cut impact on well performance (low GOR and high WC)

Figures 1.3 and 1.4 reflect additional scenarios of the effect of water cut and GOR on high GOR well performance. **Figure 1.3** shows relatively low water cut in a high GOR well with ceasing production due to high backpressure exerted across the surface pipeline at the surface. On the other hand, **Figure 1.4** demonstrates the high water cut effect on well performance where it causes an appreciable decline in oil production due to the high production of water. It also shows that high GOR has a severe impact on oil production as well. Thus, both high water cut and high GOR are severely impacting the well performance. For this specific case, a sidetrack to another reservoir might be a feasible option or equip the well with Artificial Lift method such as Electrical Submersible Pumps (ESP).

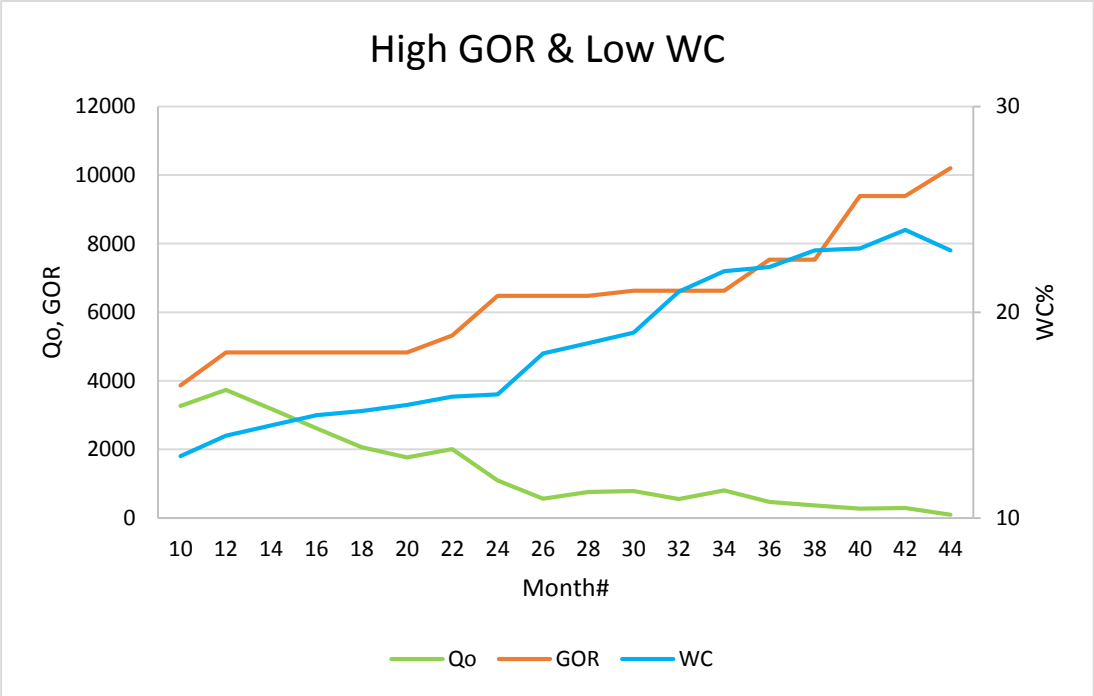


Figure 1.3 Gas oil ratio and impact of water cut on well performance (high GOR and low WC)

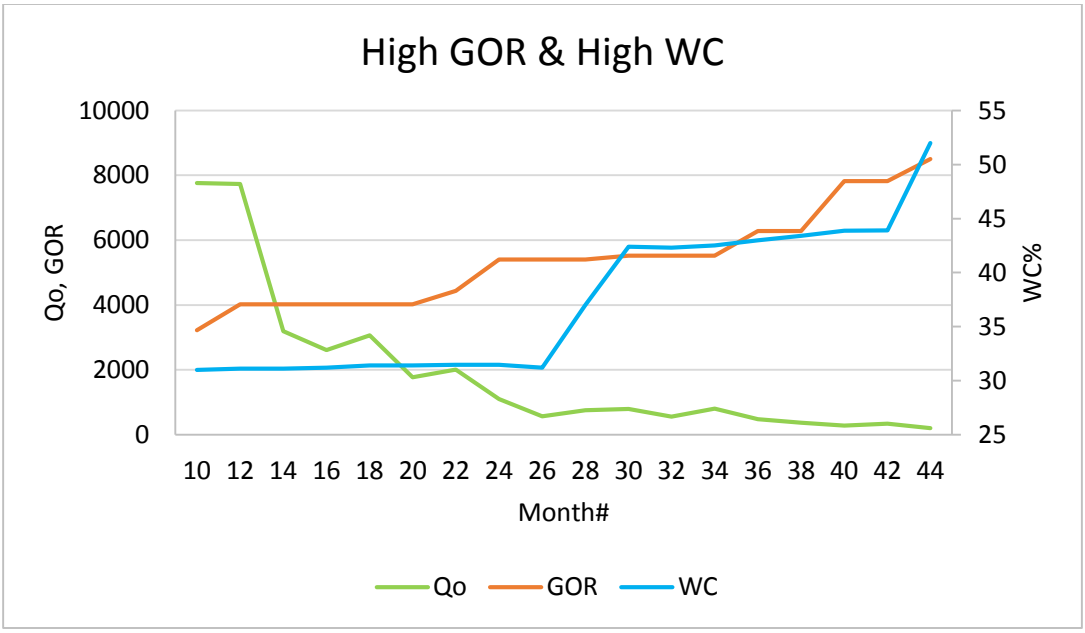


Figure 1.4 Gas oil ratio and water cut impact on well performance (high GOR and high WC)

Monitoring water-cut could lead to maximizing recovery instead of maximizing production and hence jeopardizing the health of the reservoir. There are available technologies and software to monitor water cut on a real-time basis with the objective of maximizing the recovery avoiding the need of considering secondary or tertiary enhanced oil recovery (EOR). Also, inaccurate water-cut measurement will lead to an invalid reservoir simulation model that is developed to optimize or forecast the recovery.

High water production can cause chemical corrosion in sweet/sour reservoir environments where H_2S or CO_2 are present. Chemical corrosion occurs due to the presence of soluble organic acids in the produced formation water or carbon dioxide within gas in the reservoir. These components react with the metal and cause severe pitting. Therefore, by getting an accurate water-cut value at

a high water production rate, this will allow avoidance of anticipated chemical corrosion from occurring.

II. Impact of High Water Cut and High GOR on Designing Surface Facilities

Accurate water-cut measurement will optimize CAPEX for designing or upgrading surface processing facilities in high GOR wells. Nevertheless, water cut is an essential input in the design of surface facilities at the early stage of the field. Surface facilities including water and gas handling or processing equipment are designed based on pre-defined calculations of water cut and gas-oil ratio. **Figure 1.5** shows the CAPEX as a function of water cut. As the water cut increases, more CAPEX will be spent (Open.edu, 2016). Water-cut is a crucial parameter to be considered in designing surface facilities especially when reservoir pressure is below the bubble point pressure where gas can cause significant errors in rates and water cut measurement. An inaccurate input value of water cut might underestimate or overestimate CAPEX, which might require an unexpected upgrade of the handling facilities at the later stage of field production. Therefore, accurate water-cut measurement is desirable. Also, when having high gas production, an additional cost will be added.

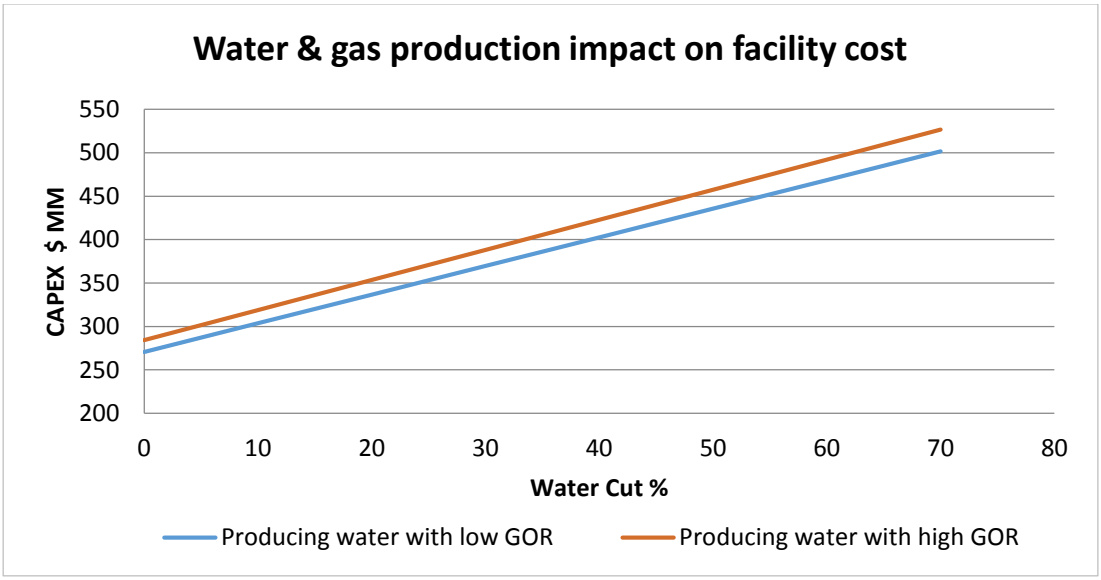


Figure 1.5 Water and gas treatment impact on facility cost (www.open.edu, 2016)

Figure 1.5 illustrates the effect of water production on the required CAPEX for processing high volumes of water and gas. Also, another observation is that much money has to be spent to treat water while a minor additional cost is required to treat gas. It is clear from the figure that CAPEX required to handle water is more than what is needed to process gas. Water processing facilities require CAPEX in low GOR wells. When having high GOR wells in association with high water production, a minor cost is added when upgrading the treatment facilities. For this reason, water-cut was considered to be the desired calculation output rather than gas because the water processing facility has more impact on CAPEX than gas processing facility. High CAPEX in association with water production confirms the importance of estimating water cut in this research.

1.3 Issues in High GOR Wells

There are several issues associated with high gas production. Below are some issues related to high gas production in high GOR wells:

I. When there is a high GOR produced at the surface, downstream flowing wellhead pressure drop increases, which lowers oil production and exerts high backpressure on the surface line as shown in **Figure 1.6**.

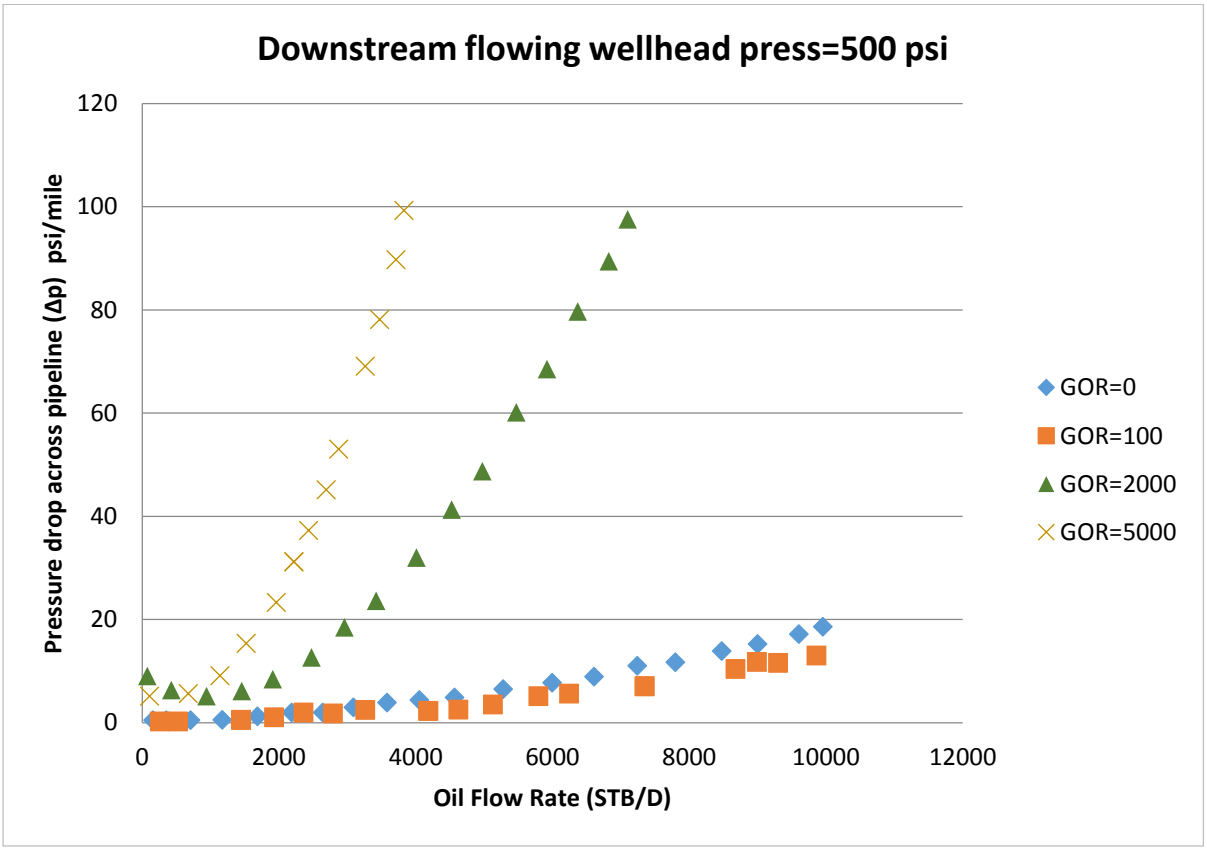


Figure 1.6 Impact of high GOR on downstream wellhead pressure drop (Moshfeghian 2014)

Figure 1.6 demonstrates the effect of high GOR on production. As GOR increases, more exerted there is more back pressure, resulting in lowering production due to additional frictional losses in the pipeline.

II. At a GOR of 3,300 scf/STB or greater (McCain, 1990), condensate might form causing wellbore plugging with liquids due to high water production, which is known as liquid loading.

Additionally, there are multiple reasons for considering water cut estimation over the gas oil ratio:

- The water-cut formula includes an oil flow rate component which is highly impacted by the presence of gas (high GOR). Thus, this affects the accuracy of water-cut estimation.

- The oil flow rate is important, but the water cut is a more critical parameter as accurate water-cut measurement is required to promote the right decision and manage unwanted water production.

CHAPTER 2 LITERATURE REVIEW

In this chapter, literature is reviewed to highlight technologies or models for measuring or estimating water cut.

I. Water Cut Metering Technologies

II. Water Cut Estimation Models

III. Multiphase Flow Correlations

2.1 Water Cut Metering Technologies

Few metering technologies are applicable to a high GOR environment. Two technologies, MFI and ABB VIS meters, were examined for high GOR wells (GOR up to 80,000 scf/STB) with high accuracy in water cut and rates measurement, and are yet to be proposed for a trial test in the subject field. Another technology was discussed, Technology of Optimization of Production (TOP), which is a pressure regulator to control the excessive production of condensate in high GOR wells. Conventional multiphase flow meters (MPFMs) are still inadequate to capture flow rates and water-cut measurements in high GOR wells. Although MPFM has a radioactive densitometer that measures relative fractions of oil, water, and gas using gamma-ray attenuations that are captured by the gamma detector but yet MPFM is unable to measure all fluid phases beyond a gas volume fraction of 90% (Nasri, A., 2014). Nevertheless, water-cut metering technologies are described below.

Hatton (1990) illustrated a microwave-based water-cut monitor technology. This device technology was developed to measure water cut using microwave signals. This tool is a cost-

effective technology where it accounts for oil composition, temperature variation, water salinity, crude properties, and water conductivity. It is a compact system and requires less footprint. Microwave signals transmitted through a reflection area from fluids at microwave frequency, as in **Figure 2.1**, are based on strong polarity which indicates that there is high permittivity due to the substantial electrical energy where water has a polarity larger than in oil. The drawback of this technology is that it will not measure water cut precisely when having free gas at the surface and also when the emulsion is taking place at the surface.

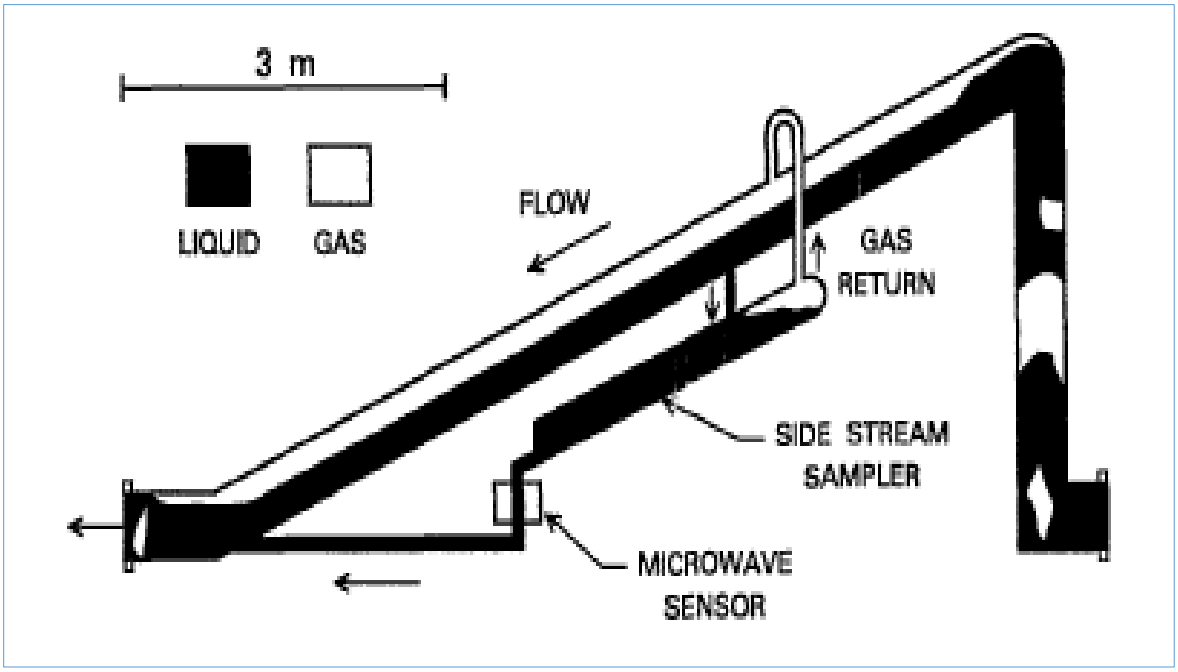


Figure 2.1 Gas-liquid setup for the microwave-based water-cut monitor (Hatton 1990)

Cellos (1999) discussed MFI multiphase meter technology which, was tested in U.S. fields. It has a separator to partially separate liquid from gas and a Coriolis flowmeter as in **Figure 2.2**. This meter was developed for high GOR applications. It could measure wells' rates from 100 to 15,000

bbl/day and with GOR up to 80,000 scf/STB. The measurement is based on a microwave technology where fractions of flow rates in the pipe are measured. The accuracy of oil rate is 5%, and the gas rate is 1-2%. This type of meter could be a good candidate to be tested in high GOR wells, but it might only be applicable to U.S. fields.

Figure 2.2 MFI multiphase meter (Cellos 1999)

Oglesby (2006) came up with a portable multiphase flow meter (MPFM) for high water-cut applications, but Gas Volume Fraction ranges from 10 to 20 % as in **Figure 2.3**. A water-cut meter is one of the meter components, which is not affected by salinity and is insensitive to low entrapped gas. It can also measure high oil flow rates and high water cut.

Figure 2.3 Portable multiphase production pester (Oglesby 2006)

Al-Saiyed (2008) and Al-Mutairi (2011) demonstrated a case study for using Red Eye metering technology. This meter is a Weatherford water cut meter, which has two components; a water cut monitoring sensor and Coriolis flowmeter as illustrated in **Figure 2.4**. The measurement is independent of density changes but applicable for low GOR up to 550 scf/STB and water cut of 5 to 42%. The infrared meter measures water cut based on near-infrared absorption spectroscopy (Al-Saiyed 2008) where water absorbs considerable electromagnetic energy. This meter provides real-time measurement. Accuracy and availability of well real-time measurement can lead to corrective action promptly on time.

Figure 2.4 Weatherford water cut meter (Al-Saiyed 2008)

Nasri (2014) multiple MPFM technologies were trial tested in high GOR wells, but none of them passed. MPFM technology has been utilized in the industry effectively for the past 20 years. There is no such MPFM operates with a high gas-oil ratio and high volume fraction of more than 95%. Three MPFMs were trial tested for multiple wells with high GOR, but none of them succeeded. Liquid rates at high gas volume fraction start to diverge due to increase in uncertainty caused by the rapid change of flow regime; mist, and annular at high gas fraction; which affects the oil rates as well.

Arsalan (2015) addressed various techniques in his paper for permanent downhole water cut measurements:

1. Relative permittivity where there is high polarity or permittivity for water.
2. Near infra red ray where water absorbs more energy at a higher wavelength.
3. Time domain transmissometry where it measures the transmission through water, oil or gas where transmission is slower in water due to high density.
4. Coriolis effect where water has a minimum oscillation frequency.
5. Gamma rays absorption for radioactive meters where produced water has higher radioactivity than oil or gas.

Nevertheless, these techniques are for the applications of black oil models where a low GOR is encountered. They are compared based on environmental impact, equipment reliability, measurement accuracy, and power requirements.

Tseytlin (2016) brought new technology into the industry for testing wells with high GOR called Technology for the Optimization of Production (TOP). This technology is a pressure regulator installed downhole along the tubing to control the condensate production for high GOR wells. This regulator controls the flowing bottom hole pressure (P_{wf}) to be at the optimum pressure at which gas is kept in the condensate and reduces the condensate flow and also maintains P_{wf} to avoid any unnecessary reduction. The range of applicability of GOR is up to 5,000 scf/STB.

Genolini (2016) exemplified the ABB VIS MPFM which is a non-gamma meter that measures individual oil, water, and gas rates for wells with high GVF of more than 90%. **Figure 2.5** shows a compact system with this metering technology where liquid and gas are measured separately as single phases with considerable accuracy with no calibration required. Overall flow rates are calculated through the sampling ratio such that the ratio between probe and pipe areas and the gas flow rate are measured separately using a Venturi meter as part of the MPFM body. The accuracy of this meter measurement can reach up to ± 5% in water cut. This meter has not been trial tested in the field. This technology is more applicable to wet gas wells.

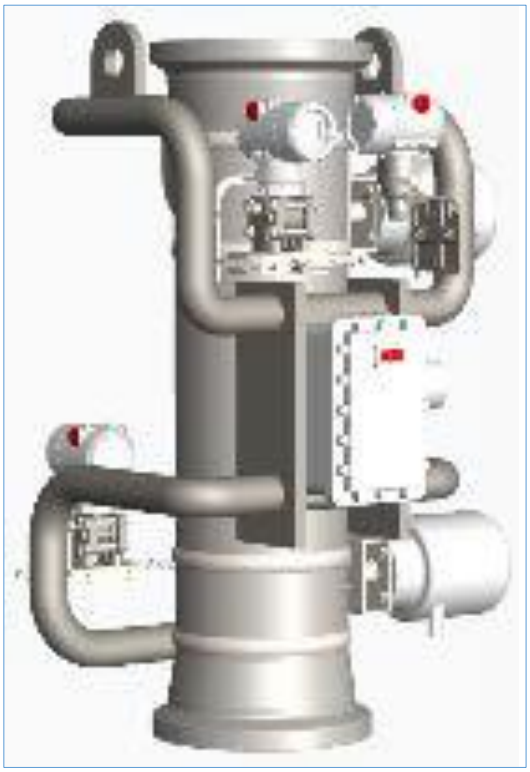


Figure 2.5 High gas MPFM ABB VIS (Genolini, 2016)

2.2 Water Cut Estimation Models

A model for classifying and assessing wells with high GOR was developed to optimize the wells' oil production better and avoid condensate production. Also, a correlation was established to calculate the gross rate combining surface and subsurface data including water cut and GOR but limited to ESP wells only. Nevertheless, most of the applications are two-phase systems where large volumes of gas are not captured. More elaboration on this part is below.

(Peruzzi 1999) based his optimization model on empirical observations and tools where examples are demonstrated in gas handling for oil wells. Wells with the average gas-oil ratio of 17,000 scf/STB were studied and optimized. Wells with high gas-oil ratio were classified into two categories; restricted (choked back) and unrestricted. Also, the challenges of excessive water and gas and solutions were explored. These models are for a specific type of wells such as ESP wells. Environmental impacts due to flaring gas were assessed when the design of gas processing facilities was underestimated.

Ghareeb (2007) developed a correlation to calculate the gross production rate at the surface considering the impact of GOR, temperature, and water cut for wells equipped with electrical submersible pumps (ESPs). Well and reservoir parameters were used to develop this new correlation, and the least squares method was utilized to find the coefficients. A disadvantage to this correlation that it is limited to artificially lifted wells and black oil reservoirs.

$$Q = \frac{9.2 \times 10^{-3} T_{th}^{3.27} H^{1.2} A^{0.81} GOR^{0.041}}{T_{bh}^{1.2} WC^{0.046}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

Re-arranging the equation results in:

$$WC = \left[\frac{9.2 \times 10^{-3} T_{th}^{3.27} H^{1.2} A^{0.81} GOR^{0.041}}{T_{bh}^{1.2} Q}\right]^{1/0.046} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

Li (2011) proposed a new model for forecasting water cut in oil reservoirs considering exponential analysis (Arps) where many variables are constant. Five models were discussed; Ershaghi-Omoregie, Liu, Warren, Purvis, and Lawal. The models with their corresponding equations are listed below:

1. ***Ershaghi-Omoregie Model***

$$\left[\frac{1}{f_w} - \ln\left(\frac{1}{f_w} - 1\right)\right] \propto N_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

2. ***Liu Model***

$$[\ln(\frac{N_p}{W_p}) - \frac{N_p}{W_p}] \propto N_p \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

3. ***Warren Model***

$$\left(\frac{N_p}{W_p + N_p}\right) \propto N_p \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5)$$

4. ***Purvis Model***

$$WOR + 1 \propto N_p \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.6)$$

5. ***Lawal Model***

$$f_{wt} = 1 - f_{ot} e^{-\alpha t} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

$$Q_t = Q_{ti} e^{-wt} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.8)$$

The models were developed to correlate water cut with the production time. These models apply to oil-water systems only where the presence of gas is negligible.

$$f_w = 1 - \frac{1}{c(1 - e^{-Dt}) + b} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.9)$$

Camilleri (2011), Al Enezi (2012), and Al Amri (2012) published three in which they developed an analytical method to estimate rates and water cut using real-time data from ESP downhole pressure and temperature sensors. The range of applicability in GOR was targeted for black oil reservoirs. Also, this method is applicable for wells with ESPs only.

Safin (2016) proposed a model to estimate water cut in horizontal wells using a statistical variogram as the best fit model between water cut and recovery factor in horizontal wells. The modeled variogram is a function of reservoir dip angle, oil column thickness, well construction and permeability anisotropy. A variogram can relate between two points by estimating the variance. The model was addressed to black oil reservoirs only.

2.3 Multiphase Flow Correlations

This section will highlight several widely used multiphase flow correlations that are used to estimate rates at the surface. Few correlations are applicable for condensate reservoirs; they are applicable to wells with low oil rates. Several multiphase (liquid/gas) flow correlations with their limitations (Production Technology 2017, Rao 1998, and Guo 2007). These correlations will be used to compare water cut with field measurement.

Poettmann-Carpenter (1952) correlation covers a limited range of liquid flow rates with gas-liquid ratios of less than 800 scf/STB.

The Fancher & Brown (1963) correlation is applicable for slim wells with tubing inner diameter of 2 3/8 and 2 7/8 inch. It is also applicable in wells with a gas-liquid ratio of less than 5,000 scf/STB and low liquid rate below 400 STB/D. This correlation is used for well model calibration.

Duns and Ros (1963) correlation is applied in wells with high gas-liquid ratio but with low liquid flow rates range.

A Hagedorn & Brown (1965) correlation is used widely in oil wells with moderate to high liquid flow rates; however, it over-estimates a pressure drop when the gas-liquid ratio exceeds 5,000 scf/STB.

Beggs & Brill (1973) correlation is used widely in oil wells. It accounts for various flow regimes and pipeline inclination, however, it also over-estimates pressure drop when the gas-liquid ratio exceeds 5,000 scf/STB.

The Gray (1978) correlation provides reasonable results in vertical/deviated wells with a GOR higher than 3,300 scf/STB.

Guo-Ghalambor (2005) covers a limited range of gas-liquid ratios with values less than 2,000 scf/STB

CHAPTER 3 PROBLEM STATEMENT AND RESEARCH OBJECTIVE

In this chapter, the problem statement and research objective are discussed:

3.1 Problem Statement

Based on the case studies demonstrated above and a literature review, the following points are considered as motives for conducting this research:

I. **Multiphase flow correlations are inadequate to estimate water cut in high GOR wells accurately. These multi-phase flow correlations are classified for low to moderate GOR wells.**

II. **Reservoir management and cost-effective field operations require another option/tool to estimate water cut in high GOR wells.**

III. **Multiphase flow metering technologies are not designed to measure water cut in high GOR wells**

Metering technologies generally fail in high GOR wells by affecting water cut measurement accuracy. Metering technologies such as multiphase flow meter (MPFM) has limitations and commonly produce errors in measurement due to improper calibration or need for maintenance rendering them useless for elongated times. Current MPFM technology provides accurate measurements for black oil reservoirs however when GOR starts to increase, MPFM accuracy becomes questionable. Another drawback of this technology is that there are limited number of manufacturers, high spare part prices, frequent calibration and maintenance, and use of radioactive sources. In high GOR wells, measurement accuracy is highly affected by the presence of gas when

gas comes out of solution near the surface of the well. MPFM measurements (Nasri, 2014) were compared to actual measurements by portable test separator and showed highly deviated values as explained in example 1 whereas example 2 shows reliable measurement from MPFM at normal conditions where GOR is less than 2,000 scf/STB. In other words, if MPFM measures a high water cut value while a well calibrated portable testing separator shows a zero or low water cut value, this will result in a wrong decision when depending on MPFM measurements and so will cost unnecessary workover rig expenses if water shut off job required. Therefore, it will be a considerable addition to the industry when coming out with a new methodology that fills that gap. This gap necessitates the need for developing an alternative method or correlation that could estimate rates and water cut within the acceptable ranges. Two examples are demonstrated where the first example shows metering equipment performance for black oil reservoir conditions where the latter showing inaccurate measurement at abnormal conditions (Nasri, 2014).

Example-1; at Black oil reservoir conditions at GOR < 2,000 scf/STB

MPFM measures flow rates in all phases, water cut and GOR with considerable accuracy when GOR is less than 2,000 scf/STB. **Figures 3.1 and 3.2** compare between conventional test separator and MPFM readings for gas oil ratio and water cut, respectively. Both GOR and WC are falling within the acceptable error band of ± 10% against a calibrated portable test separator measurement. Therefore, MPFMs provide operationally acceptable field measurement in low GOR wells.

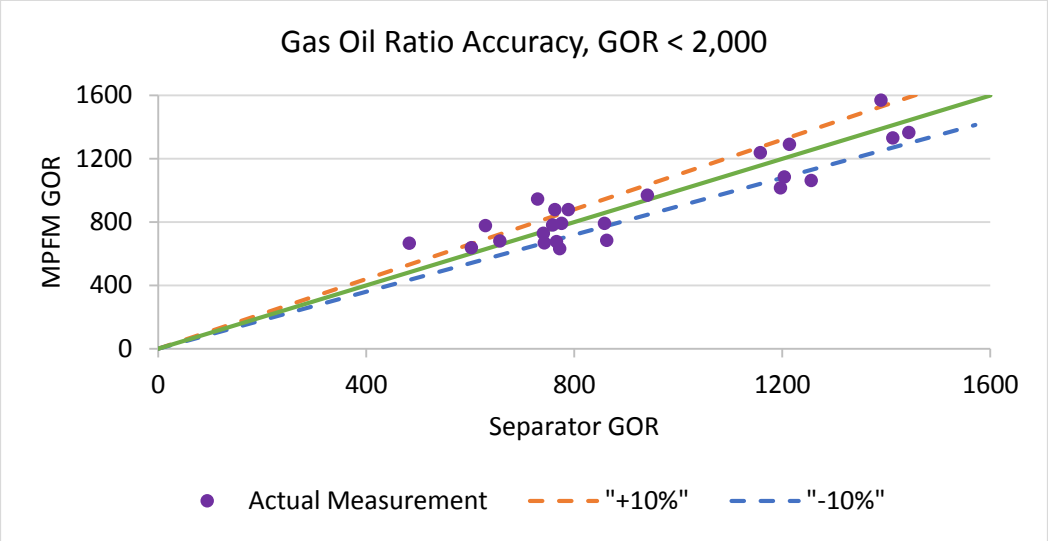


Figure 3.1 Comparison between separator and MPFM GORs less than 2,000 scf/STB

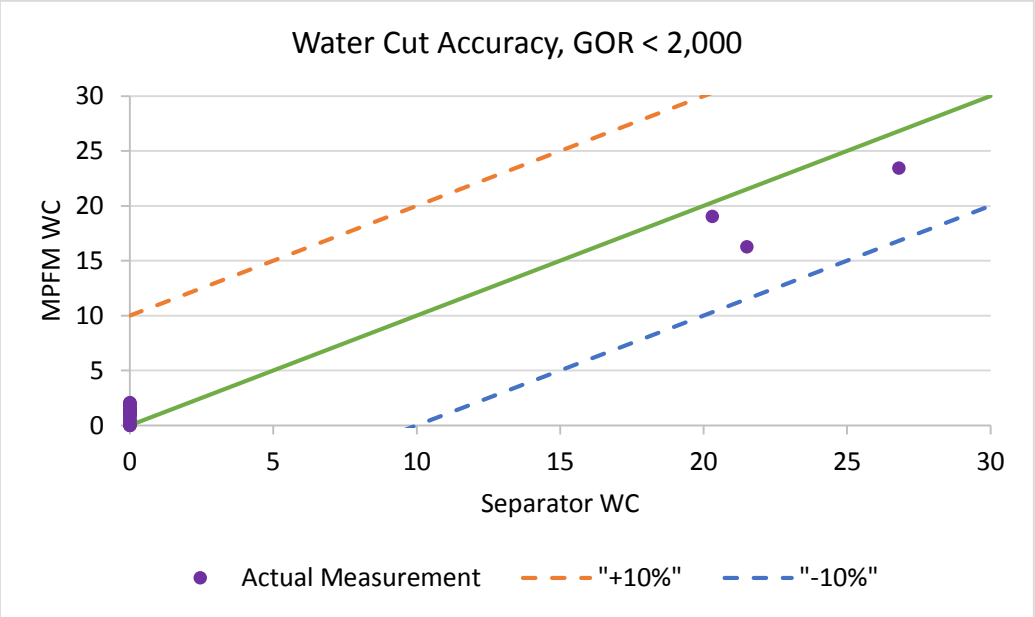


Figure 3.2 Comparison between separator and MPFM WCs for GOR less than 2,000 scf/STB

Example-2; few metering technologies qualified at a GOR >= 2,000 scf/STB

In contrast, wells with GOR greater than 2,000 scf/STB were tested using both test separator and MPFM. At high GOR (>=2,000 scf/STB), the deviation between MPFM GOR and test separator GOR was more than ± 10% in terms of absolute error as **Figure 3.3**. Nevertheless, this is also impacting the WC readings, where MPFM is failing to measure the WC accurately at high GOR as **Figure 3.4**. Therefore, high GOR affects rates measurement and water cut values. Therefore, this confirms that there are no certified MPFMs for high GOR (Nasri, 2014). Additionally, water-cut values from MPFM in **Figure 3.4** indicate that most of the wells are wet producers whereas the separator measurements confirm the opposite. The improper decision making will lead production engineers to take incorrect decisions about well performance issues. Therefore, MPFMs produce a deviation in both GOR and WC in high GOR wells.

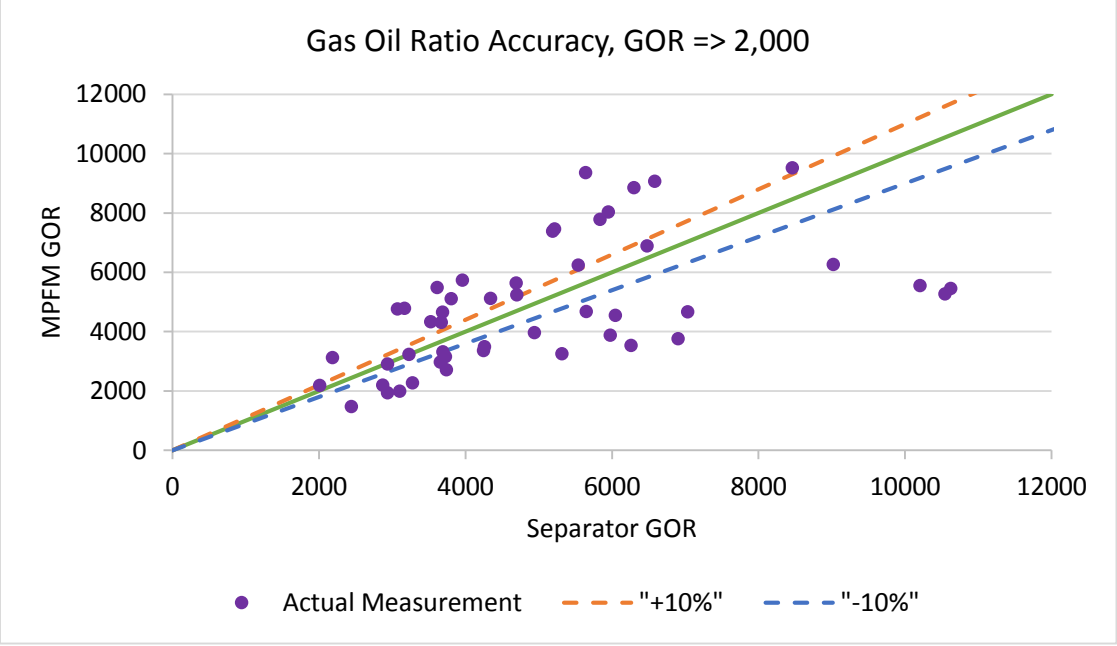


Figure 3.3 Comparison between separator and MPFM GORs >= 2,000 scf/STB

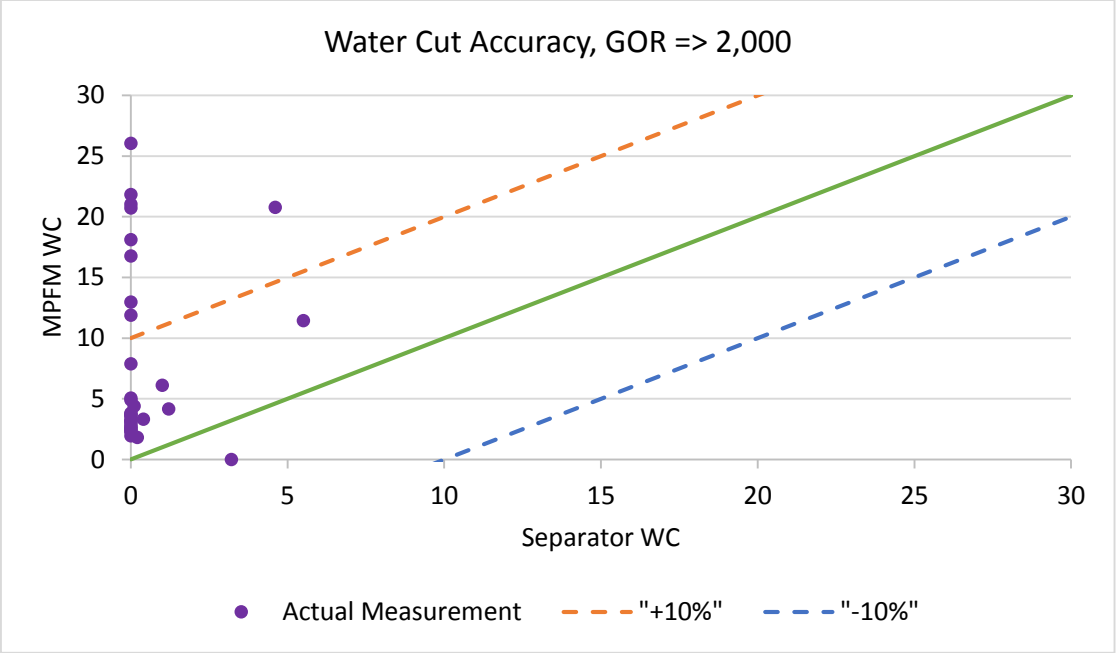


Figure 3.4 Comparison between separator and MPFM WC values for GOR >= 2,000 scf/STB

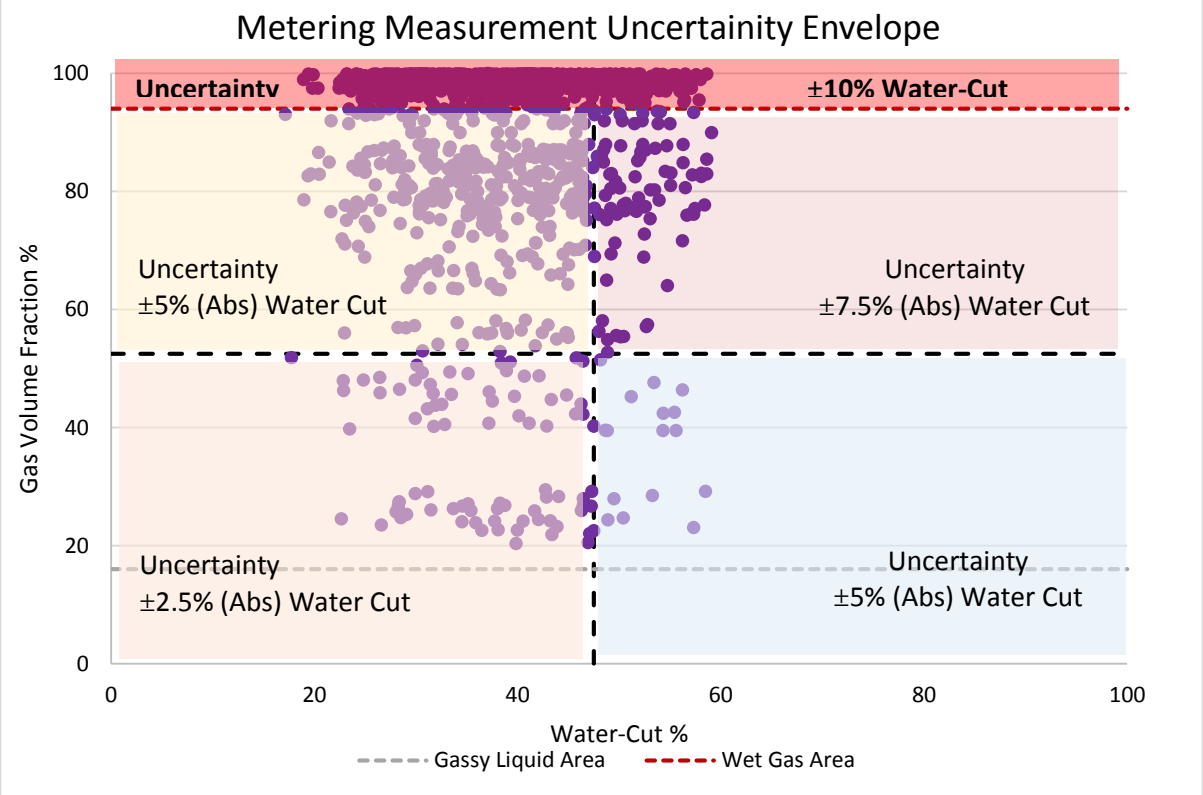


Figure 3.5 Standard metering envelope (Handbook of Multiphase Flow Metering, 2005)

In **Figure 3.5**, the metering envelope of GVF against WC (both in percentages) consists of three regions which are the gassy liquid area, wet gas area, and moderate WC and GVF area. The GVF is defined as the ratio of gas flow rate to the total flow rate. In this research, the majority of wells are lying inside or near the wet gas area where the GVF is higher than 90% and WC ranges from 10 to 70% with an uncertainty that reaches up to 10%. In other words, for high GOR or GVF wells, the error in liquid rates, hence in water rate, will deteriorate causing high uncertainty in metering measurement with an error of 10% in water-cut that could be tolerable ('Handbook of Multiphase Metering' by NFOGM, 2005). Conventional MPFM can provide accurate measurements for wells with GVF up to 90% whereas, beyond that, measurement accuracy issues will dominate causing errors in liquid and so in water rate measurement.

3.2 Research Objective

The primary objective of this research is to develop an improved model (s) to estimate water cut accurately in high GOR wells using surface and subsurface field data to achieve at least one of the benefits below:

- Avail a methodology to reduce the dependency on portable test separators and MPFMs that can be a cheaper option to better estimate water-cut in a diverse environment.
- Provide another means to measure devices besides test separators as they can be used when a test separator is unavailable or out of service.
- Save the cost of trial testing metering technologies.
- Enable effective reservoir management due to accurate water-cut estimation in high GOR wells.

CHAPTER 4 MODELING TECHNIQUES

Under this chapter, an introduction to two models will be discussed and used in this research:

- **Non-linear Multiple Regression Model**
- **Artificial Neural Network (ANN) Model**

4.1 Nonlinear Multiple Regression

Regression theory acquires analysis of independent variables to construct a predictive model that is mapped to the dependent variable. The main aim of regression analysis is to identify the relationship between independent and dependent variables that result with the least square of the errors (Ray, 2015). There are two main categories in regression modeling: Linear and non-linear regression. Linear regression requires a linear relationship between inputs and output where there is a constant multiplied by or added to the input with each input of power 1. For non-linear regression applications, it is used to identify the nonlinear relationship between inputs and output with any form of data. That is, the non-linear regression produces a nonlinear model for understanding the behavior and relationship data and achieving a better output. It is also worth noting that the coefficient of determination (R^2) can be considered as a measure of how good the nonlinear relationship is between inputs and outputs.

There are two types of non-linear regression. The first type is the simple non-linear regression where there is one input and one output exhibiting a non-linear relationship. With the second type, which will be used in this research, there are multiple inputs and one output displaying a non-linear

relationship. In this case, an empirical correlation or ANN is used to quantify the relationship between various inputs with one single output.

- **Difference between correlation coefficient and coefficient of determination:**

The correlation coefficient (CC) is a measure of the degree of variation or proportionality between two parameters (MathBits, 2018). A CC with a negative value means that there is an inverse relationship between inputs and output. CC with positive values explains a direct relationship, and zero CC indicates no relationship between independent and dependent variables. A CC studies the relationship between one input and output at a time demonstrating the type of relationship, whether direct or inverse.

The coefficient of determination (R^2) is defined as the linear relationship between input and output (MathBits, 2018). R^2 with a high value means there is a strong relationship between inputs and output. R^2 with low values explains a weak relationship and zero R^2 indicates no relationship between the independent and dependent variables. A coefficient of determination reflects the percentage deviation in the output with respect to all inputs where a higher value of R^2 will be preferable. A coefficient of determination is applicable to both simple linear and multiple regression models where it dictates the relationship between multiple inputs with respect to an output (Bansal, 2015). Additionally, **Figure 4.1** demonstrates the effect of coefficient of determination R^2 on reducing the error. As the relationship between inputs and output tends to become stronger, the bigger the value of R^2, the lesser the error in getting predicted values compared with the actual values with high accuracy.

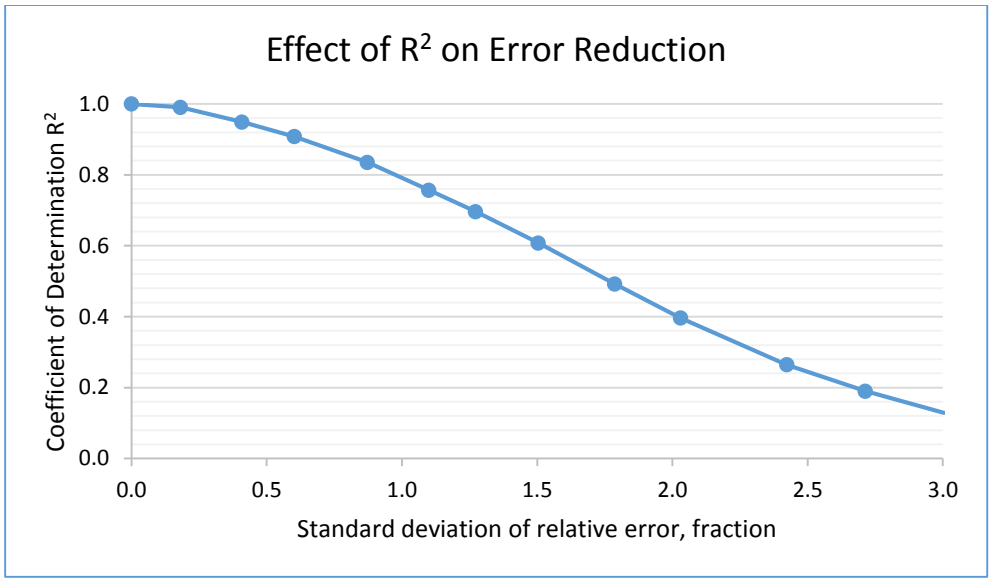


Figure 4.1 Effect of R^2 on error reduction (Lake, 2003)

4.2 Artificial Neural Network (ANN)

Over the past two decades, Artificial Neural Networks (ANNs) started to peak up although they did not change the world drastically at that time. That is not because they are not useful for prediction but the technology was insufficient to facilitate neural networks where there is a vast amount of data that require high processing computers. An ANN is considered one of the most popular deep learning techniques. The objective for neural networks is that they should mimic some aspects of how the human brain operates but this objective has not yet been fully achieved. In the human brain, 100s of billions of neurons are connected with language capabilities, unlike computers, which need to be fed with inputs to solve a specific problem. The brain of the neural network is modeled through hidden layer (s) where many neurons in hidden layers are interconnected together, and input values are processed through hidden layers until getting the

desired target (Udemy, 2018). **Figure 4.2** addresses the three layers with inputs and output for a fully connected neural network with lines demonstrating the weights (w) and biases (b).

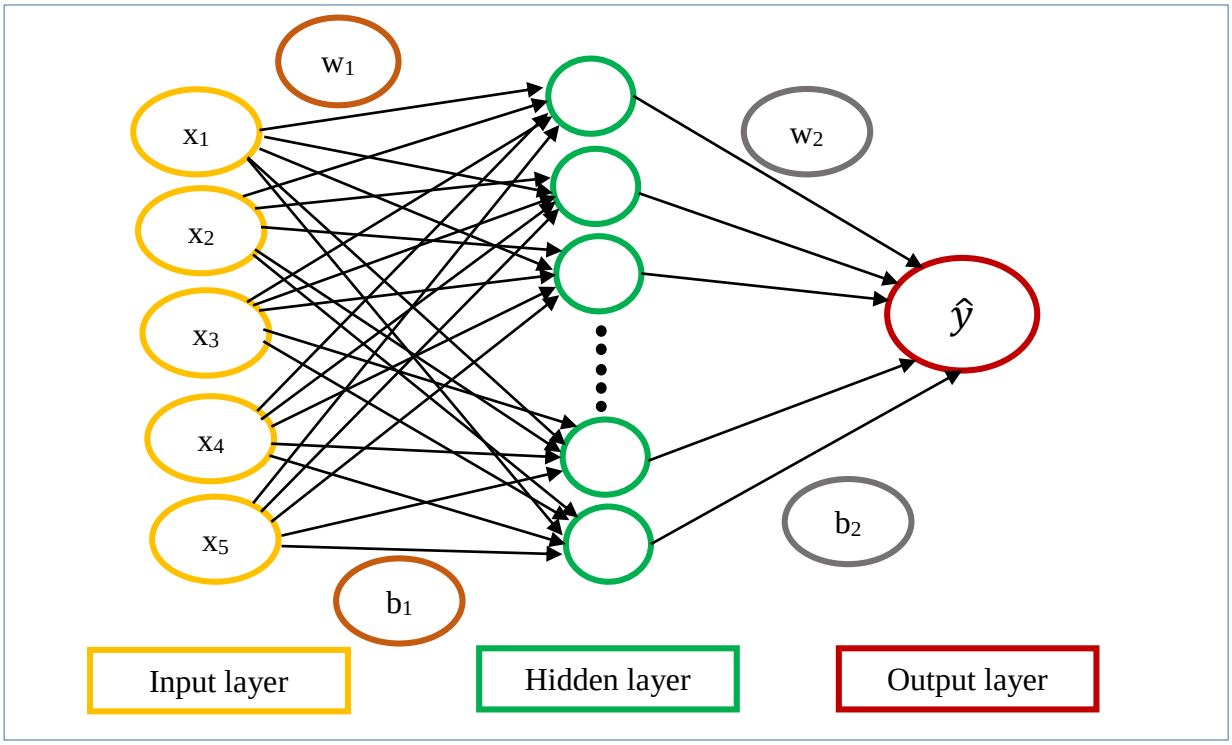


Figure 4.2 Simple fully connected neural network structure (Udemy, 2018)

Neural Network (NN) Components

A neural network consists of neurons, layers, and activation (transfer) function.

- **Neuron:**

Neural network layers contain many neurons. The neuron is the primary building block of an ANN. It is the essence of ANNs to mimic how the human brain operates. The neuron consists of the components listed below as per **Figure 4.3** (Farouk, 2013):

- Dendrites; receiver of the signal for a neuron

- Axon; the transmitter of the signal for a neuron
- Dendrites of other neurons are connected to an axon
- An axon does not touch dendrites
- Synapse; connections or lines linking each layer; along which signals pass

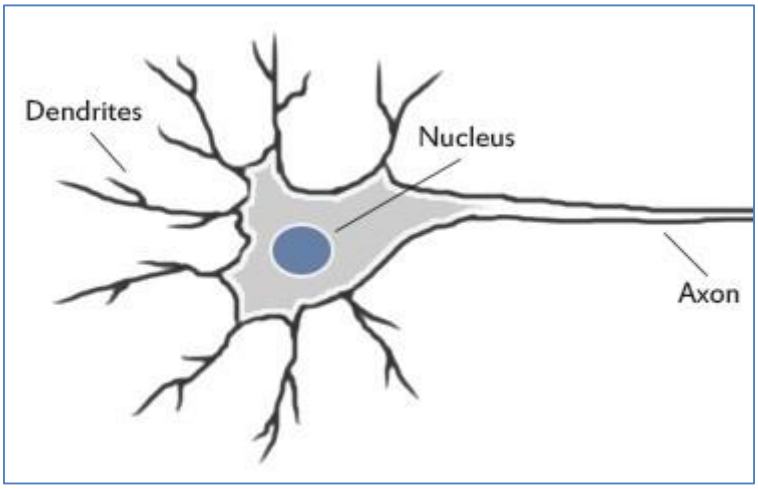


Figure 4.3 Human neuron components (Pokharna, 2016)

What happens inside the neuron?

Figure 4.4 illustrates the NN components where weighted values are added, and the transfer function passes signals from one neuron to another via synapses as follows:

1. Step 1: Weighted values are added

$$\sum_{i=1}^{m} w_i \, x_i$$

2. Step 2: Transfer function; to pass on signal or not (from one neuron to another)

$$\emptyset\left(\sum_{i=1}^{m} w_i\, x_i\right)$$

3. Step 3: A signal passes along the neuron (synapses) to another when the transfer function is applied and weights, representing the connections, are adjusted.

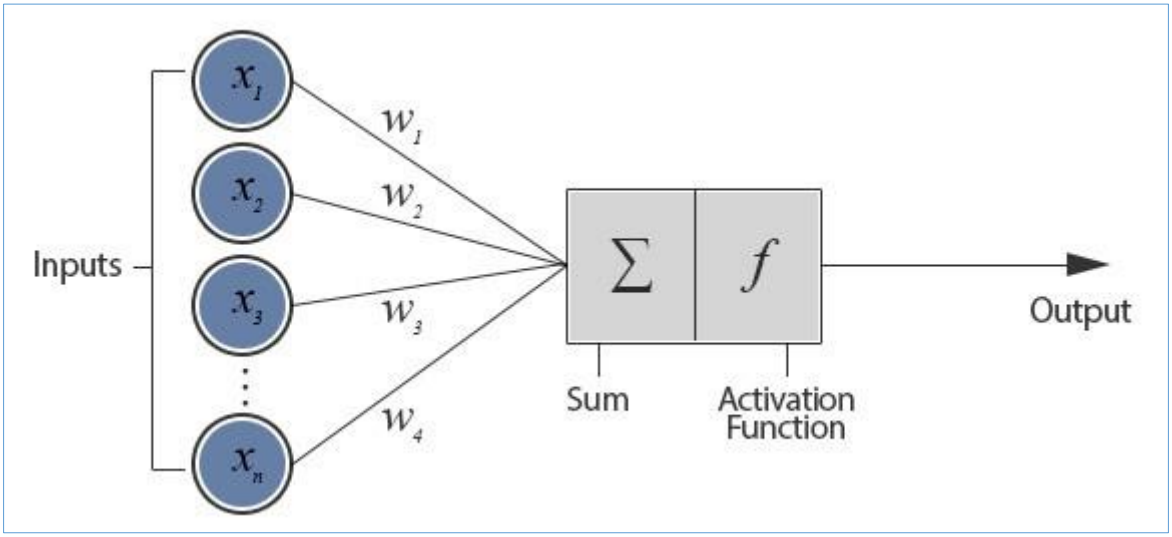


Figure 4.4 Artificial neuron model (Pokharna, 2016)

Neurons in the network are learning via adjusting or modifying weights. Some weight values might have zero values and others will have non-zero values. For instance, five inputs exist in the input layer in the neural network, namely, oil flow rate (Q_o), upstream flowing wellhead pressure (P_{wh}), downstream flowing wellhead pressure (P_{ds}), Gas Oil Ratio (GOR) and flowing bottomhole pressure (P_{wf}) with one output which is water cut (WC). The hidden layer consists of three neurons. The first neuron in the hidden layer might pick inputs Qo and GOR since their weight values are non-zero to be the most important variables affecting WC. As the oil flow rate increases, water cut might increase due to sweep efficiency. Also, increasing the GOR will make oil shrink, and so the

oil flow rate will vanish with time, and water production will rise. The second neuron in the hidden layer might pick one input which is flowing bottomhole pressure as it has a non-zero value while remaining inputs are zeros. As flowing bottomhole pressure rises, this will drag up the oil-water contact level to expand to make water cut increase. The third neuron in the hidden layer has three non-zero inputs which are upstream flowing wellhead pressure, downstream flowing wellhead pressure and oil flow rate. When increasing upstream flowing wellhead pressure and reducing downstream flowing wellhead pressure, this will increase the oil flow rate, which might eventually increase water cut in return. Some neurons leverage the importance of a specific number of inputs to the output. Also, another aspect is that weights might be positive or negative, which it will impact the relationship between input and output based on the sign of the weight. The neural network might pick up things we would not consider in the first place. That is why an ANN is considered to be one aspect Artificial Intelligence (AI) techniques, which finds the relationship between irrelevant parameters.

- **Layers**

A neural network is a group of layers with nodes in each layer. Nodes are considered as an input for the next layer. A neural network comprises of three main layers:

i. Input Layer:

In this layer, the count of neurons is equal to the number of inputs.

ii. Hidden Layer (s):

The input layer is connected to the hidden layer (s) by synapses where the transfer function is activating the weights and passing on them to the hidden layer. One hidden layer is usually

sufficient to carry out complex problems. In the hidden layer, a non-linear transfer function is used for the output to become a linear separable solution. A linear separable solution contains neurons that can be separated with a single decision line or surface which enables proper learning (H. Pokharna, 2016).

iii. Output Layer:

Once the weights are passed on from the input layer to the hidden layer, the output layer receives the weight where the inputs multiply them. If the output value is not equal to the target, then the weights are adjusted.

- **Weights and Biases**

Weights in NNs are considered to be the most important component. Weight is defined as how strong is the connection between the input to the hidden layer node where it indicates the relationship between input and output. It helps in determining how strongly each neuron affect each other. Weights are generated when the data training starts and they can be negative or positive. Negative weights consider an inverse relationship between input and output. The bias value will enable shifting the transfer function either to the left or to the right. The matrix form the equation for the NN model follows:

$$\mathbf{w}_1^T \mathbf{p} + b_1 = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (4.1)$$

w_1: Weights vector between input and the hidden layer of the neural network.

& w_2: Weights vector between the hidden and the output layer of the neural network.

b_1: Bias vector between input and the hidden layer of the neural network.

& b_2: Bias between the hidden and the output layer of the neural network.

The general equation that combines all components of NN:

$$Output = \sum weights * inputs + bias \quad \text{..... (4.2)}$$

In **Figure 4.5**, the output is plotted against input where three different weights were examined for the sigmoid transfer function. It is noticeable that changing the weight value is changing the smoothness of the transfer function. As the weight increases, the curve less steep or flatten.

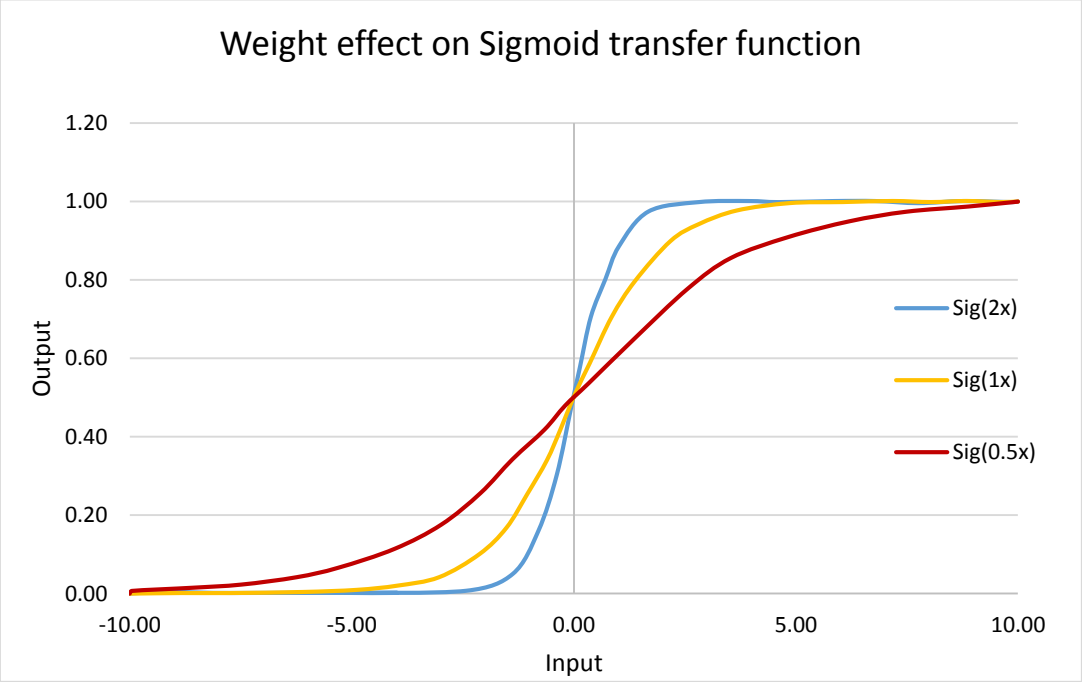


Figure 4.5 Weight effect on sigmoid transfer function (stackoverflow website)

In **Figure 4.6**, the output is also plotted against input where three different biases were studied for the sigmoid transfer function. It is evident that changing the bias value is shifting the curve to the right or left where a weight of -5 for w_1 shifts the curve to the right side.

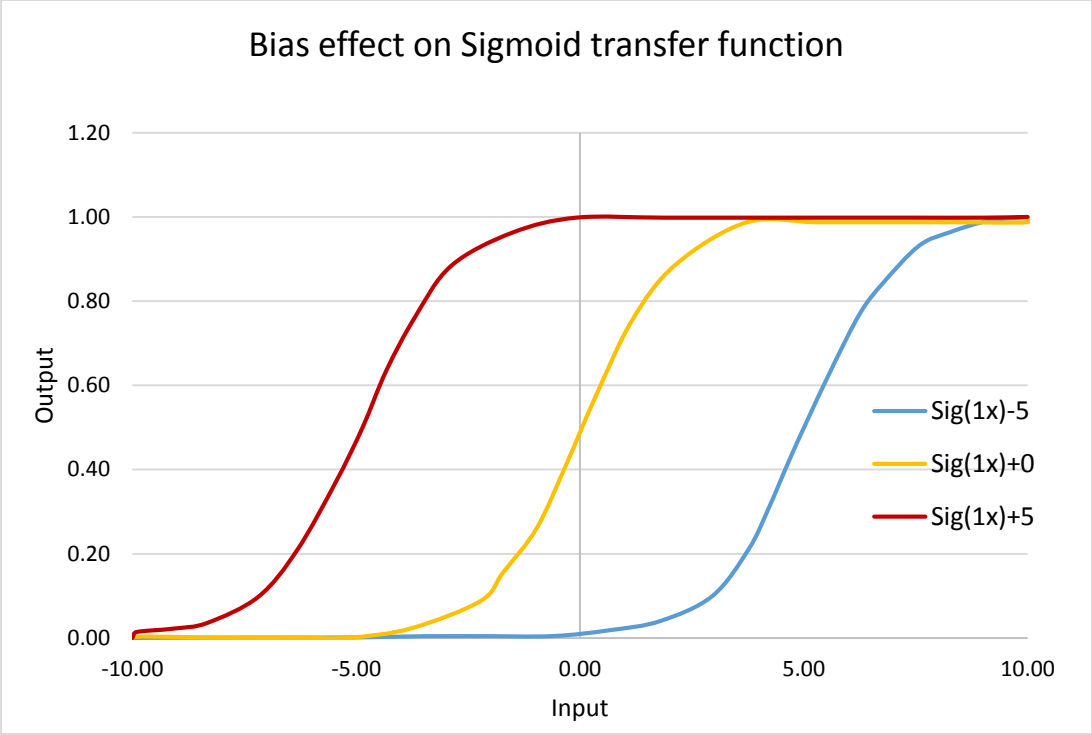


Figure 4.6 Bias effect on sigmoid transfer function (stackoverflow website)

- **Back Propagation Algorithm**

It consists of two processes:

I. Forward propagation process: Information is entered into input layers, and then it is propagated forward to get output values and compared to actual values. Then errors are backpropagated through the network in the opposite direction where network training is achieved by adjusting the weights.

II. The backpropagation process is an advanced algorithm allows us to adjust the weights at the same time. This algorithm is well structured simultaneously (which weights to adjust).

- **Training (Learning) functions**

ANNs learn by training the network by modifying weights and biases. The training algorithm is classified into three main types (Farouk, 2013):

i. **Supervised Learning:** This consists of multiple inputs and one output. It utilizes all available inputs to predict the output by adjusting weights and biases.
ii. **Unsupervised Learning:** It has inputs only where weights and biases are adjusted subject to these inputs.
iii. **Reinforcement Learning:** Not as popular as supervised learning. It has inputs and scores outputs to check if the predicted output is matching with actual.

Backpropagation is a general algorithm that is utilized when performing learning algorithms such as gradient descent optimization and minimization of the cost function (Nielsen, M. 2015). **Table 4.1** lists some key training algorithms that are accompanied by the backpropagation method. Errors during the training process are subject to forward and back propagation where forward propagation results from neurons of input and hidden layers being activated and weights being updated to predict the output. In reverse, back propagation entails errors are backpropagated when output is still no match with the target and weights are updated accordingly.

Table 4.1 Typical Training (and Testing) Functions used in ANN

Type#	Type Description	MATLAB function
0	Levenberg-Marquardt with one hidden layer	trainlm
1	Levenberg-Marquardt with two hidden layers	trainlm
2	Backpropagation using gradient descent	traingd
3	Backpropagation using gradient descent with momentum and adaptive (variable) learning rate backpropagation	traingdx
4	Backpropagation using Broyden, Fletcher, Goldfarb, and Shanno (BFGS) Quasi-Newton	trainbfg
5	Backpropagation using scaled conjugate gradient	trainscg
6	Backpropagation with One Step Secant	trainoss
7	Resilient backpropagation	trainrp
8	Batch backpropagation	trainb

Like any other prediction tool, the backpropagation algorithm may exhibit solution convergence issues. There are several training functions with the essence to modify the general backpropagation algorithm as listed in **Table 4.1**. Below are features of some vital training algorithms to better predict the target with minimal drawbacks:

- The Levenberg-Marquardt training function is considered to be the fastest algorithm in terms of convergence to the target value with achieving highly accurate results, but the accuracy in this algorithm might decrease when the number of inputs and their weights increase as it will require more storage requirements. It is used mainly with NN but not for pattern recognition.
- The gradient descent training function has weights and biases that are updated after each iteration or after completing all iterations. The gradient is calculated based on the change in weights and biases however this process is time-consuming and might slow the training process.
- Variable learning rate gradient descent training function has a learning rate added to the training process. The selection of a proper learning rate is a crucial factor to enable the success of the training process. Setting the learning rate to be too large or too small might affect the solution convergence.
- The Conjugate Gradient algorithm is an advanced technique that is used with a large number of inputs and weights achieving faster convergence.
- The One Step Secant function is based on the Quasi-Newton algorithm. The advantage of this function is to optimize the computational time and storage requirements.
- The resilient backpropagation function is used in pattern recognition problems.
- The BFGS quasi-Newton backpropagation function achieves faster optimization but requires much storage requirements when there is an extensive network.

The NN training is performed with the following steps:

Step 1: Randomly initialize the weight to small numbers close to zero (but not zero) until the error is minimized.

Step 2: Input the first point of the dataset in the input layer; each feature in one input route.

Step 3: Forward propagation: From left to right, the neurons are activated in a way that the weights limit the impact of each neuron. Then, propagate the activations until achieving the predicted result $\hat{y}$.

Step 4: Compare the predicted result to actual. Calculate the error.

Step 5: Backpropagation: From right to left, the error is backpropagated. Update the weights accordingly to how much they come responsible for the error. The learning rate decides by how much the weights are updated.

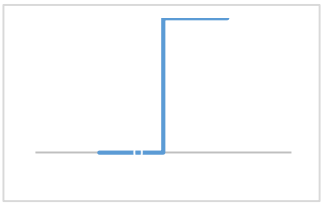
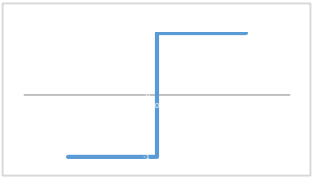
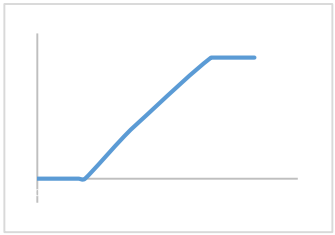
Step 6: Repeat steps 1→5 and update weights for each observation (reinforcement learning) or repeat steps 1 → 5 but update the weights only after a batch of observations (batch learning).

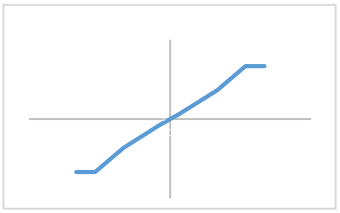
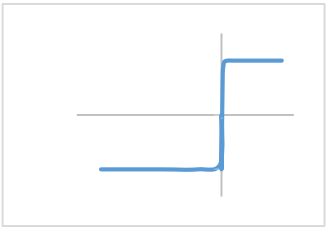
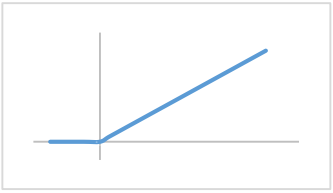
Step 7: When the whole training set passed through ANN which makes an epoch. Re-do more epochs.

- **Transfer functions Ø(x)**

The transfer function passes on the signal from one neuron to another and applies to the output. There are commonly used transfer function in either hidden or output layers, as in **Table 4.2**:

Table 4.2 Typical Transfer functions (Farouk, 2013)

Function Name	Input / Output Mathematical Relation	Input / Output Graphical Relation	MATLAB Function
Hard Limit	$\begin{cases} y = 0 & x < 0 \\ y = 1 & x \geq 0 \end{cases}$		hardlim
Symmetric Hard Limit	$\begin{cases} y = -1 & x < 0 \\ y = 1 & x \geq 0 \end{cases}$		hardlims
Linear	$y = x$		purelin
Saturating Linear	$\begin{cases} y = 0 & x < 0 \\ y = x & 0 \leq x \leq 1 \\ y = 1 & x > 1 \end{cases}$		satlin

Function Name	Input / Output Mathematical Relation	Input / Output Graphical Relation	MATLAB Function
Symmetric Saturating Linear	$\begin{cases} y = -1 & x < -1 \\ y = x & -1 \le x \le 1 \\ y = 1 & x > 1 \end{cases}$		satlins
Log-Sigmoid	$y = logsig(x) = \frac{1}{1 + e^{-x}}$		logsig
Hyperbolic Tangent Sigmoid	$y = tansig(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} = \frac{2}{1 + e^{-2x}} - 1$		tansig
Positive Linear	$\begin{cases} y = 0 & x < 0 \\ y = x & 0 \le x \end{cases}$		poslin

Below are some points of comparison between different transfer functions:

- The linear transfer function is used when there is no need to apply any threshold to estimate the output.

- The sigmoid transfer function is the most accessible tool in terms of differentiation since the function is smooth and continuous. Also, it effectively performs the backpropagation algorithm and produces accurate results. The output value is ranged between 0 and 1.
- Hyperbolic sigmoid transfer function follows the same characteristics as of sigmoid function, but it returns values of a range between -1 and 1.
- When inputs in the sigmoid function have enormous negative values, this will result in a zero value which might produce inaccurate zero weights and disturb the NN model.
- The symmetric saturating linear transfer function is similar to hyperbolic sigmoid transfer function as it returns values between -1 and 1. The latter is smoother than the first which performs more efficiently with normally distributed data.
- The hard limit transfer function is discontinuous and therefore non-differentiable. Therefore, the intent to use this function might be impractical for backpropagation techniques.

- **Why use ANN in Petroleum Engineering?**

ANNs have commonly been used in petroleum engineering applications for the last few years in all sort of sectors, including reservoir, production, drilling, fluid properties, well logging and many other applications (Attia, 2015). The frequent use of ANNs is due to the simplicity of deploying the ANN technique with well-trained data. Well parameters such as pressure, temperature, flow rates, and gas-oil ratio tend to exhibit a viable relationship with respect to each other which makes neurons function more efficiently in understanding the relationship between specific input with the output. Also, NNs can interlink between two independent variables with the output. For

instance, NNs can correlate between flowing wellhead pressure and impact of temperature on liquid flow rate, where there is a direct or inverse relationship, and assign a negative weight for the inverse relationship and positive weight for the direct relationship. Therefore, NNs will produce accurate results and fewer errors for a set of inputs that exhibit a physical relationship to the output. It is worth mentioning that ANN applications are not limited to rate prediction only but have been used to predict flow regimes, temperature, pressure profiles, and multiphase flow metering.

CHAPTER 5 METHODOLOGY

In this chapter, the problem approach is addressed in this research in three stages which are reviewing available multiphase flow correlations, developing a new empirical correlation and developing an ANN model. The details follow.

The steps below summarize the methodology in a nutshell:

1. **Gather data for low and high GOR wells**
2. **Perform data description and analysis**
3. **Evaluate available multiphase flow correlations to estimate water-cut indirectly**
4. **Develop two models; if necessary**
    i. Non-linear multiple regression model
    ii. ANN model
5. **Select the most applicable model that achieves the least error and cover the whole range of GOR.**
6. **Verify the optimized model with a new set of data.**

5.1 Available Multiphase Flow Correlations Evaluation

Reference to Table 2.1 in section 2.3, Five correlations are selected to be assessed against the actual water cut from the test separator to study their applicability for high GOR applications. The selected correlations are Guo Ghalambor, Gray, Hagedorn Brown, Duns & Ros, and Beggs & Brill based on their GOR applicability and limitations although they are limited to a specific range of GOR. The multiphase flow correlations will be used to indirectly estimate water cut and quantify errors in high GOR wells in a field where data is available. Multiphase flow correlations are examined for sensitivity analysis run using nodal analysis.

The calculations and error analysis are performed to estimate the water cut using multiphase flow correlations with the following steps:

1. Gather data from 100 wells (1 data point/each) considering the full range of GOR.
2. Construct well models using PROSPER^{TM} software and calibrate them accordingly in reference to the latest rate tests from the test separator.
3. Run "system" calculations by inputting P_{wh}, water cut, GOR and estimate liquid flow rate (Q_L) with using an initial water cut value from the latest rate test.
4. Iterate water cut until reaching an error of ±10% in the liquid flow rate between calculated and measured liquid flow rates.
5. Report the final water cut and compare with the actual.
6. Observe the estimated error in water cut to the actual for low GOR wells.

5.2 New empirical correlation using non-linear regression

Multivariate nonlinear regression is used to develop a model to better estimate the water cut with the following steps:

1. Gather data for 100 wells with 4/5 data points per well for low GOR wells and 19/20 data points for high GOR wells.
2. Based on the Random Forest technique results, consider five inputs to develop the new empirical correlation.
3. Group the five parameters into three with aligning these parameters to order of magnitude concept as dividing two parameters or taking the log of a parameter should result in a number between (0 to 0.1), (0.1 to 1) or (1 to 10).
4. Construct multiple combinations of the grouped inputs.
5. Report the error of each combination, if the relative absolute error is less than 10% and absolute error is less than 5%, stop.
6. If it can meet the 10% error band, consider the data points for wells with GOR of less than 2,000 scf/STB. If the relative absolute error is less than 10% and absolute error is less than 5%, stop.
7. Document the correlation for wells with GOR of less than 2,000 scf/STB.

5.3 ANN Model

An ANN model is developed to capture the full range of GOR. The steps are as follows:

1. Gather data for 100 wells with four to five data points per well for low GOR wells and 19-20 data points for high GOR wells.
2. Based on the Random Forest technique results, consider five inputs to develop the new empirical correlation. No parameter grouping will be performed to inputs as they will be fed to the ANN model.
3. Divide the data set into 70% for training and 30% for testing and validation. Use the default model setting for ANN by using the followings:

    - One hidden layer with ten neurons
    - The tangent sigmoidal transfer function
    - The Levenberg Marquardt training function

4. Report the model when the absolute relative error is less than 10%, and the absolute error is less than 5% and then stop.
5. Perform training optimization as follows:

    i. Optimize different training functions and report the one with least errors
    ii. Optimize different transfer functions and report the one with least errors
    iii. Optimize the neuron count in the hidden layer from 1 to 20 and then report the neurons counts with least errors.

6. Based on the training optimization, report the best model with the five parameters.
7. After reaching an acceptable error, utilize the nine inputs and run the ANN model.

8. If the reported errors are less than acceptable error, report the model with the nine parameters.
9. Report the final equation with weights and biases for both cases.
10. Convert the empirical equation to the normalized form as a function of actual value with the following steps:

i. Utilize **equation 5.1** to normalize inputs and outputs between -1 and 1 as follows:

$$\frac{y - y_{min}}{y_{max} - y_{min}} = \frac{x - x_{min}}{x_{max} - x_{min}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5.1)$$

Where y is the output/target and x is the input

The Hyperbolic Tangent Sigmoid Transfer function is considered as an example, y values range from -1 to 1. Therefore, **equation 5.1** can be written in the following form:

$$y = 2 * \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) - 1 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5.2)$$

ii. Use **equation 5.3** by plugging in all weights and biases to calculate the output:

$$y = [\sum_{i=1}^{N} w_{2i} * transferfunction(w_{1i,j}x_j + b_{1i})] + b_2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5.3)$$

iii. Convert the values of y, de-normalize the output using **equation 5.4**:

$$y_{\,calc} = (slope\ estimate) * y_{\,target} + (Y - int.)\ estimate \dots\dots\dots (5.4)$$

CHAPTER 6 RESULTS AND DISCUSSION

In this chapter, data are thoroughly described, and models are developed to estimate water cut better. Also, data were handled with various cleansing techniques. Selective inputs were chosen with the highest importance to the output. Results for both the nonlinear regression model and ANN are discussed. The ANN model is optimized to select the most accurate scenario.

6.1 Data Description and Handling

6.1.1 Data Description

Data required to estimate water cut includes surface and subsurface data as follows:

1. Dynamic data:
    - Gas-oil ratio (GOR)
    - Oil flow rate (Q_o)
    - Upstream flowing wellhead pressure (P_{wh})
    - Downstream flowing wellhead pressure (P_{ds})
    - Upstream flowing wellhead temperature (T_{wh})
    - Flowing bottomhole pressure (P_{wf})
    - Flowing bottomhole temperature (T_{wf})
    - Choke valve position (ranges from 0 to 100% of 3-inch choke diameter; “0%” for closed and “100%” for fully connected)
    - Inflow control valve (ICV) position (ranges from 0 to 10; “0” for closed and “10” for fully connected)

2. Static data used to generate well models:
    - PVT data such as oil formation volume factor, API, water, oil and gas densities
    - Reservoir data such as permeability, pressure, temperature, skin … etc.
    - Geometry represented by tubing depth, the hole diameter, and inclination angle

Initially, more than 5,000 data points were collected, but several cleansing techniques were performed to achieve a total of 1,210 data points. The data handling techniques were as follows:

- Outliers that are below or above the range were removed. When a data point was irrelevant to a group of data, it was called an outlier which can be defined as a value below the first quartile or above the third quartile.
- Negative, duplicated and frozen data were also removed. This type of data issues is resulted from equipment malfunctioning which requires calibration or replacement. Also, frozen data might be due to interruptions in data transmission since the data are taken from real-time devices with no missing data within the 1,210 data points.
- Conducted normality test using the skewness. The 5,000 data distribution was away from normality. A skewness of zero means that the distribution is normal with one set of data.
- The coefficient of determination (R^2) was also used between each input and output at a time. Enhancing R^2 in preparation for the ANN model will help in reducing the error as per Figure 4.1.

Descriptive statistical analysis for nine inputs and one output was performed to identify the statistical relationship between inputs and the output **in Table 6.1**:

Table 6.1 Statistical description of data used in the research

Variable	Choke size %	P_{wh}, psi	P_{ds}, psi	T_{wh}, °F	Q_o, STB/D	GOR, scf/STB	P_{wf}, psi	T_{wf}, °F	ICV Position (number)	Water cut %
Minimum	40	523	306	125	936	418	1,686	168	3	17
Maximum	100	1,656	1,251	199	11,128	10,042	2,260	242	10	59
Mean	70	1,022	584	161	4,471	5,246	1,973	204	6	39
Standard deviation	17.4	263	147.3	21.8	2029.1	2,776	21.8	22	2.3	8.8
Skewness	-0.04	0.548	0.994	0.058	0.657	0.012	-0.026	0.058	0.080	0.069
Correlation Coefficient	-0.04	-0.412	0.3151	0.0009	-0.8870	-0.7371	-0.412	0.0009	0.0313	

One hundred (100) wells with a total of 1,210 data points were included in this research representing 48 wells with a GOR of less than 2,000 scf/STB and 52 wells with a GOR of greater or equal to 2,000 scf/STB. **Figure 6.1** dictates the distribution of water cut as a normal distribution after data cleansing.

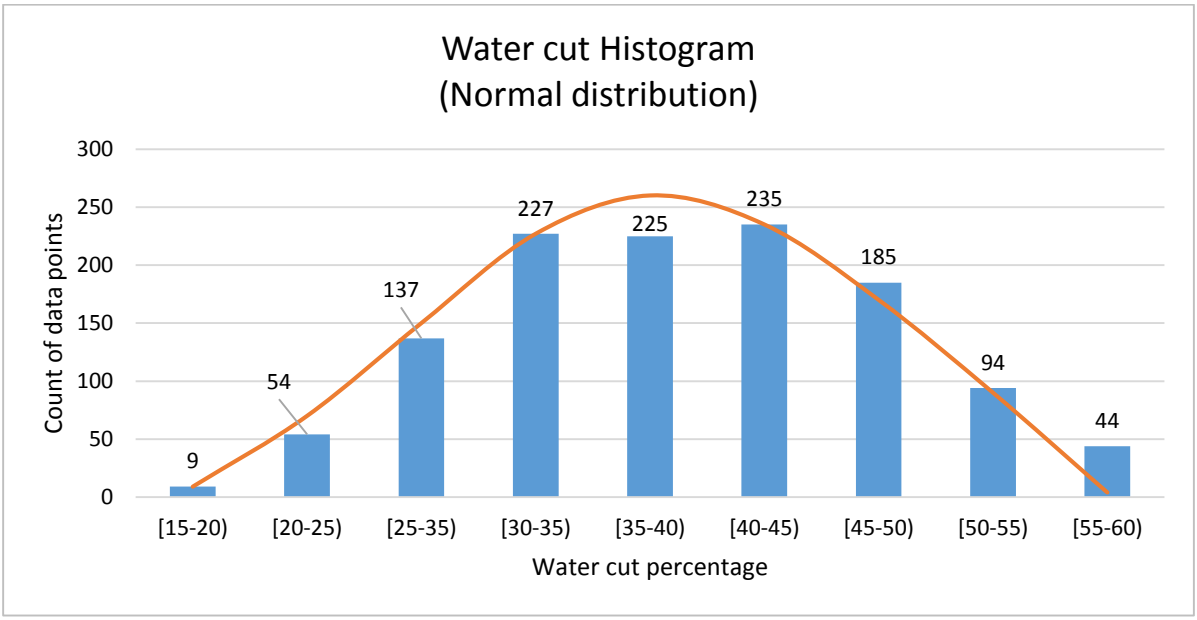


Figure 6.1 Water cut histogram

6.1.2 Basic Statistical Description Equations

- ***Mean*** is defined as the summation of all data points of a specific parameter divided by the number of data points.

$$\mu = \frac{\sum x}{N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6.1)$$

- ***Variance*** is the measure of how much data are far from the mean.

$$\sigma^2 = \frac{1}{N}\sum(x-\mu)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6.2)$$

- ***Standard Deviation*** can be calculated by taking the square root of the variance.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6.3)$$

- ***Root Mean Square Error (RMSE):*** Measures the error deviation around the mean:

$$RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}{E_t}^2} . \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6.4)$$

- ***Skewness*** is the measure of asymmetry with the subject to the normality in a distribution. If the value is close to the mean, then the skewness will approach to zero indicating a normal distribution.

$$S = \sqrt{n}\frac{\sum(x-\mu)^3}{(\sum(x-\mu)^2)^{3/2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6.5)$$

- ***The Correlation Coefficient (CC) or (r)*** is a measure of linearity in a relationship between two variables; x and y. The resultant value of CC ranges from -1 to +1.

$$CC = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sqrt{x^2} - \left(\sqrt{x}\right)^2][n\sqrt{y^2} - \left(\sqrt{y}\right)^2]}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6.6)$$

Where y is the water-cut and x is any of the input variables.

- ***The Coefficient of Determination (R^2)*** is another measure of linearity in a relationship between two variables, x and y and it is simply representing the square of the correlation coefficient. The resultant value of R^2 ranges from 0 to 1.

$$(R^2) = (CC)^2$$

- ***Average Absolute Error; E_{aa}*** is the difference between actual value and calculated value with averaging all data points.

$$E_{aa} = \frac{\sum|Actual\ y - Calculated\ y|}{N} \quad X\ 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6.7)$$

- ***Average Relative Percentage Error; E_{ar}*** % is the difference between actual value and calculated value to the actual value in percentage with averaging all data points.

$$E_{ar}\% = \frac{1}{N}\sum \frac{|Actual\ y - Calc.y|}{Actual\ y}\ X\ 100 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6.8)$$

When having a normal distribution, this will assist in reducing the standard deviation which means that all data points are located around the mean and hence decreasing errors. Also, the skewness is another measure of normality. The coefficient of determination indicates the relationship strength between actual and estimated values.

6.1.3 Sensitivity Analysis of inputs with respect to the output

The physical relationship between multiple inputs to output can be quantified by conducting a sensitivity analysis is used to define how input values are affecting output value under specific assumptions. In this research, water-cut is the desired output that is impacted by high GOR. The liquid rate is also affected by the GOR as well. **Figures 6.2 to 6.7** demonstrate the fundamental relationships between each independent and dependent variable in reference to the data described in section 6.1.1.

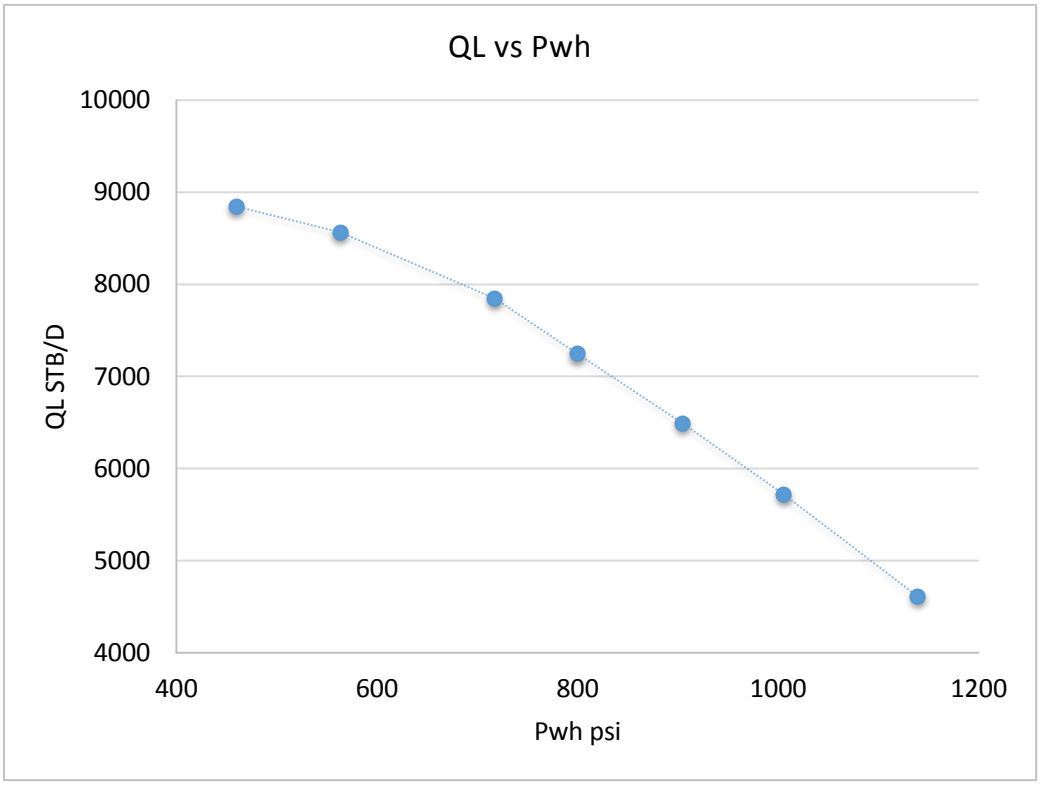


Figure 6.2 Relationship between liquid flow rate and flowing wellhead pressure

Figure 6.2 shows the inverse relationship between flowing wellhead pressure and liquid flow rate. It is noticeable that as the upstream flowing wellhead pressure increases, the liquid flow rate decreases at the surface.

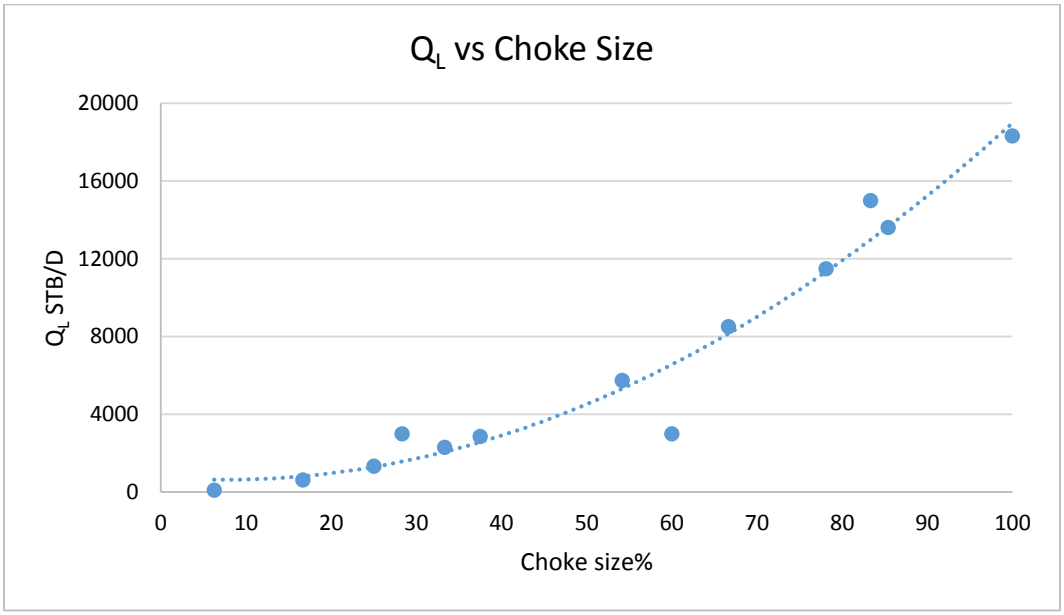


Figure 6.3 Relationship between liquid flow rate and choke size

As shown in **Figure 6.3**, as the choke size increases, the liquid flow rate increases since the choke is the rate controller at the surface.

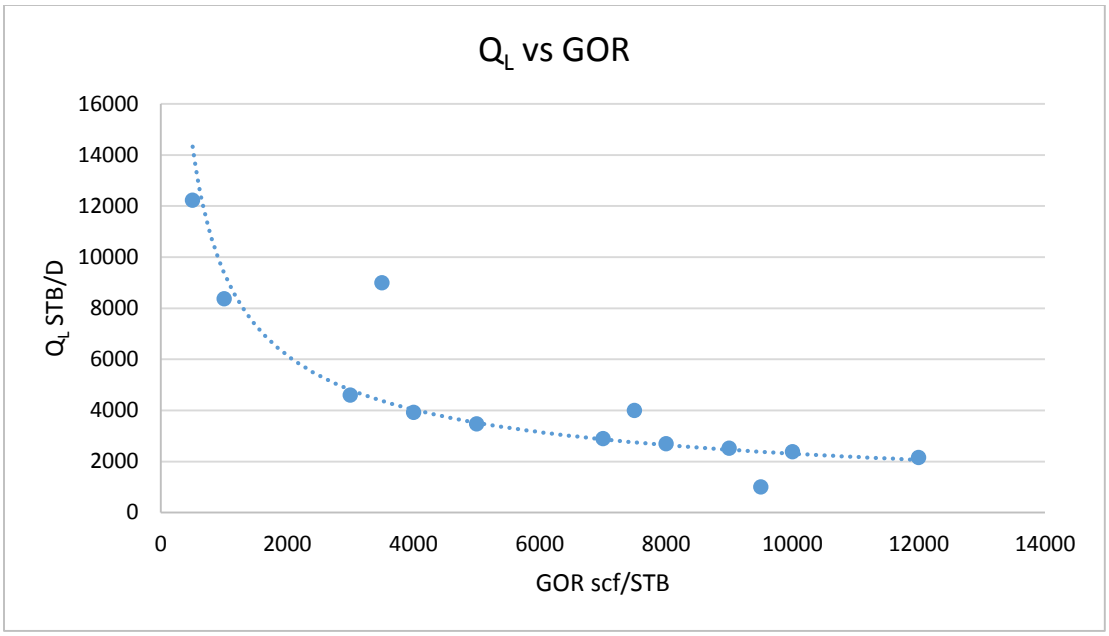


Figure 6.4 Relationship between liquid flow rate and gas oil ratio

As per **Figure 6.4**, the liquid flow rate is affected by a high GOR whereas the GOR increases, the liquid flow rate becomes less as the gas occupies more volume against the oil flow rate.

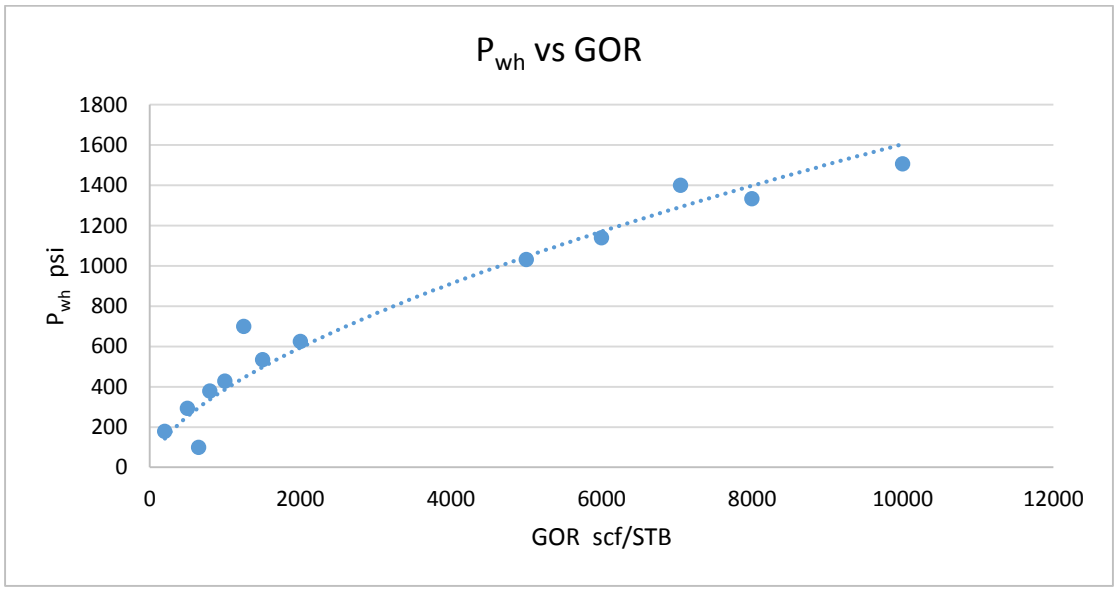


Figure 6.5 Relationship between flowing wellhead pressure and gas oil ratio

The GOR has a direct impact on flowing wellhead pressure. As the GOR increases at the surface, the flowing wellhead pressure increases as shown in **Figure 6.5.**

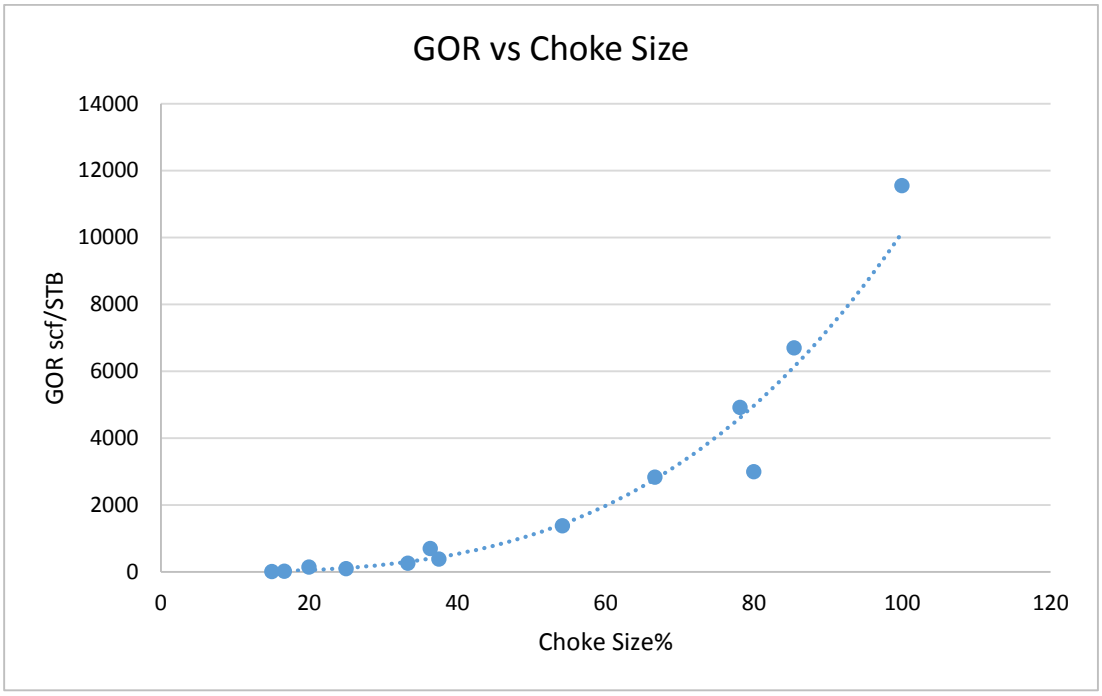


Figure 6.6 Relationship between GOR and choke size

Figure 6.6 illustrates the direct relationship between the GOR and choke size. In this research, a high GOR condition is dominating the described data-set, which makes the GOR have a more significant effect on all dependent variables such as water cut and liquid rate. As the choke size increases production, the GOR increases accordingly.

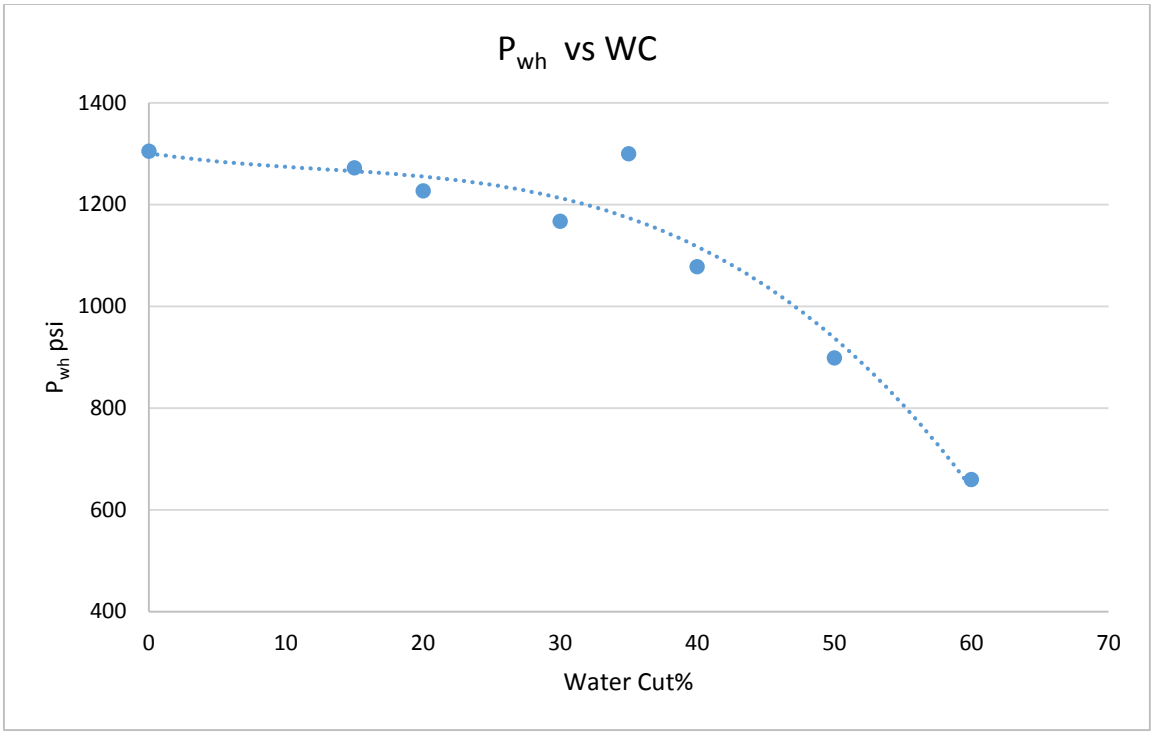


Figure 6.7 Relationship between flowing wellhead pressure and water cut

From **Figure 6.7,** as the flowing wellhead pressure decreases due to choking back the well, the water cut will eventually increase.

6.1.4 Inputs Importance

In linear and non-linear relationships, correlation coefficient and coefficient of determination are commonly used to identify the relationship between inputs and outputs knowing the exact relationship or fitting function between the two variables. When the relationship is difficult to be quantified, another measure of a relationship is used to understand better data which is called Random Forest (RF).

Random Forest

Random Forest (RF) is one of machine learning techniques used to solve complex problems. It is mainly used in classification and regression problems, which produce accurate results. RF technique randomly forms a forest. This forest contains multiple decision trees where each one creates a model with utilizing the dataset. Each decision tree compares one or multiple inputs in relations to the output. Then, based on all constructed models, the model with the highest accuracy is selected. The conventional decision tree is usually used when having one input and one output in the dataset at which this is not applicable to the area of research in this work.

Nevertheless, one of the byproducts of this technique is to identify the relative importance of each input or independent variable to the output without the need of knowing the type of fitting function. It also performs in a better mode when there is a massive amount of data points. The concept behind RF is that there is a forest with growing random decision trees where each decision tree provides a vote for each input to the desired output. Then, RF assigns weights in terms of standard error in the form of z-score for inputs with respect with the output based on their non-linear relationship (Breiman 2018). Additionally, RF dictates the relative importance by observing at how much each tree improve the clarity of the path across all other trees in the random forest. It computes z-score for each input after training and scales the results so that the sum of all importance is equal to 1.

Figure 6.8 is the result of Random Forest technique with oil flow rate (Q_o) ranked as the highest important input to the water cut:

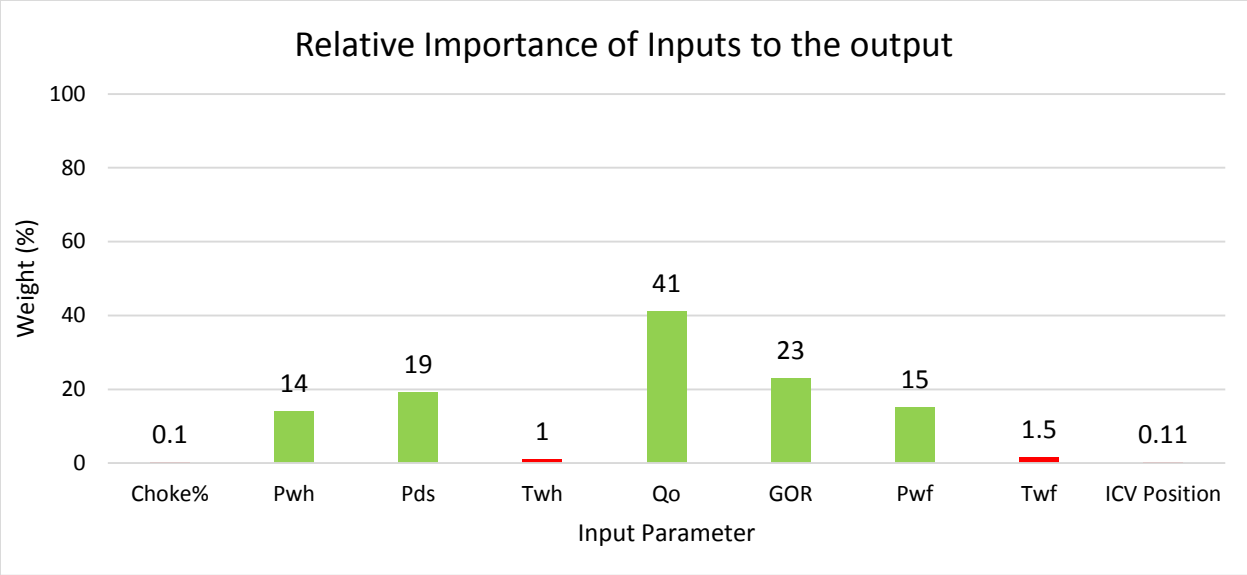


Figure 6.8 Random Forest results for ranking inputs to the output

The result of this technique that five inputs are listed to be important to the water cut which are, oil flow rate (Q_o), downstream flowing wellhead pressure (P_{ds}), Gas Oil Ratio (GOR), upstream flowing wellhead pressure (P_{wh}), and flowing bottomhole pressure (P_{wf}). These five inputs will be used in developing the new models.

6.2 Results from Available Multiphase Flow Correlations Evaluation

Reference to section 2.3, five multiphase flow correlations were selected based on GOR applicability to indirectly estimate water cut and be compared with actual water cut. The selected correlations are Guo Ghalambor, Gray, Hagedorn Brown, Duns & Ros, and Beggs & Brill. Several runs were made to come up with IPR and TPR model for 100 wells where both curves intersect at the operating conditions with high GOR. Results are presented for a full range of GOR and GOR of less than 2,000 scf/STB. **Figures 6.9, 6.12, 6.15, 6.18, and 6.21** illustrate the cross-plot of water

cut values from test separator and estimated from each correlation for all wells with a full range of GOR. **Figures 6.10, 6.13, 6.16, 6.19, and 6.22** show the cross plot of water cut for wells with GOR of greater than or equal to 2,000 scf/STB and **Figures 6.11, 6.14, 6.17, 6.20, and 6.23** reflect the cross-plot of water cut for wells with GOR less than 2,000 scf/STB. The average absolute relative error (E_{ar}) for wells with GOR of less than 2,000 scf/STB exhibits a value of 13% viewing liberal values compared to actual water cut. Average E_{ar} for wells with a full range of GOR and high GOR wells exceeds 30%, which shows the inaccuracy of these correlations to estimate water cut, which is due to shortcomings and limitations of these correlations.

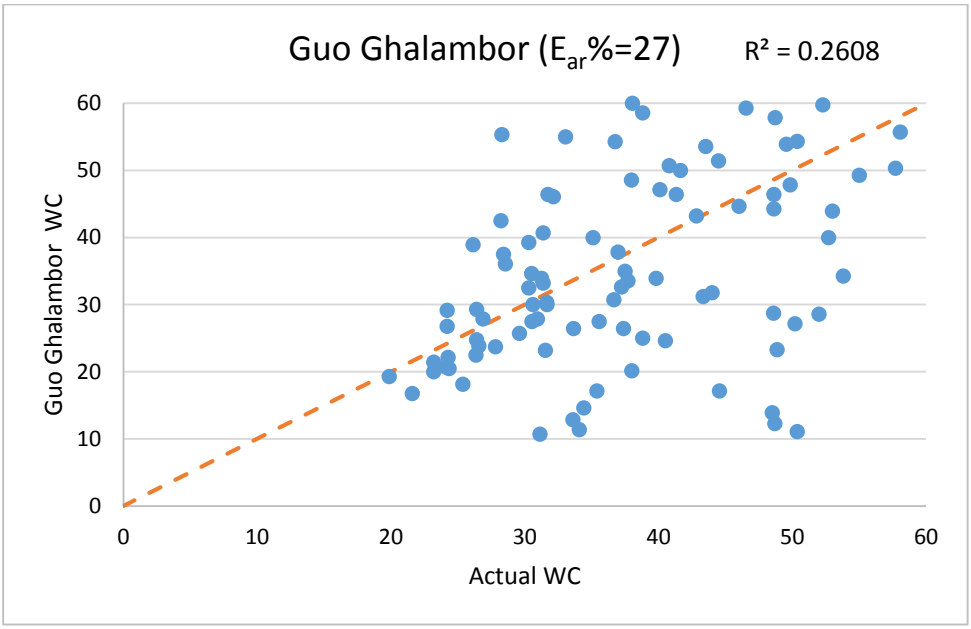


Figure 6.9 Guo Ghalambor WC comparison (all wells)

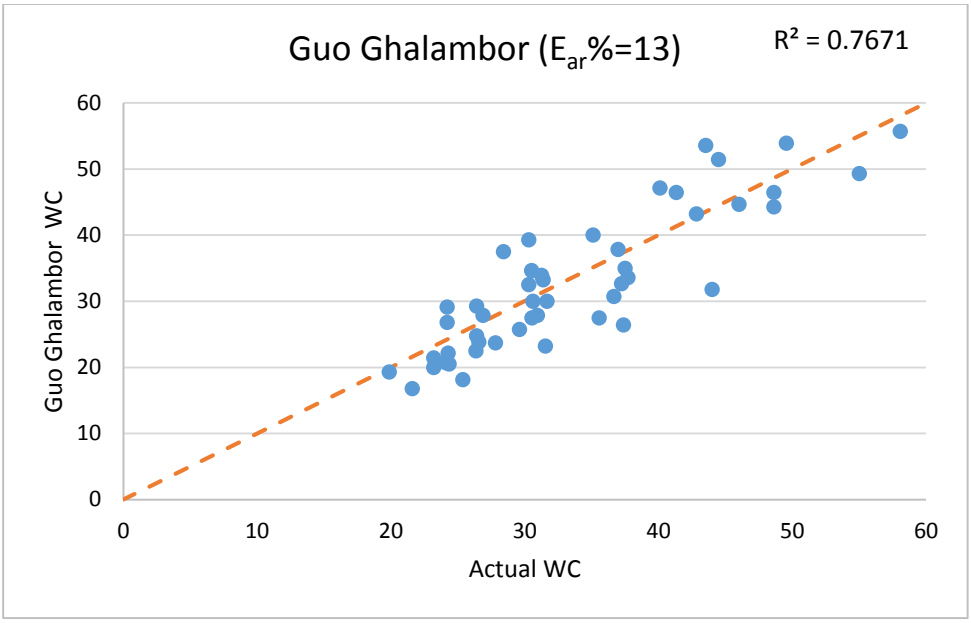


Figure 6.10 Guo Ghalambor WC comparison (GOR <2,000 scf/STB)

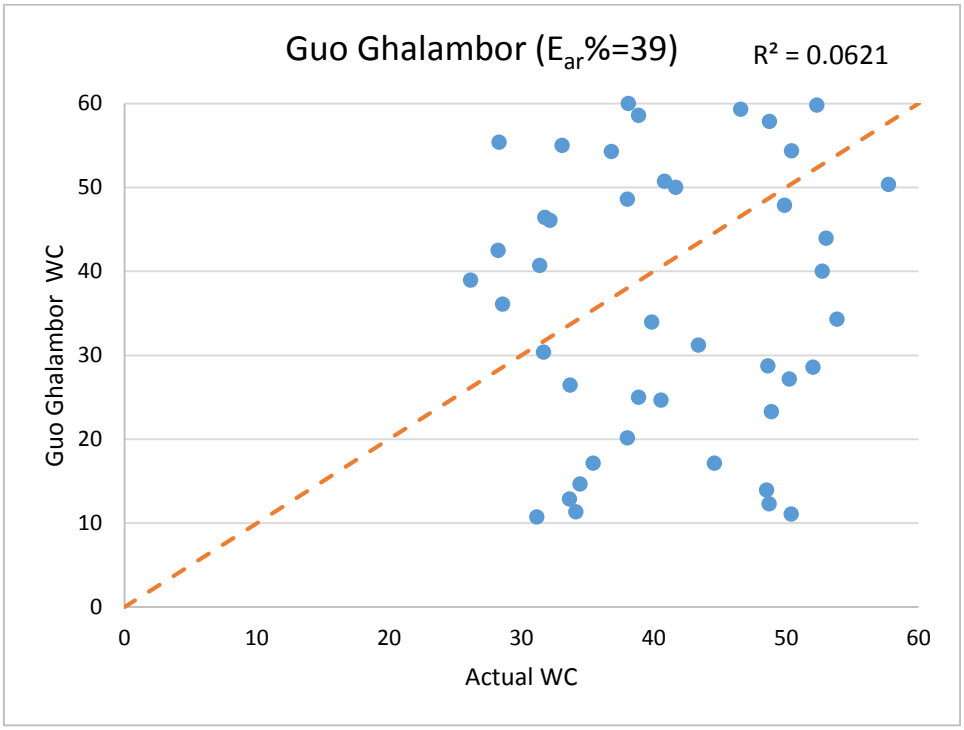


Figure 6.11 Guo Ghalambor WC comparison (GOR >= 2,000 scf/STB)

Figures 6.9, 6.10 and 6.11 reflect the calculated water cut from Guo Ghalambor multiphase flow correlation against the actual water cut noting that the average absolute relative error (E_{ar}%) of 13% when applying the correlation in wells with (GOR < 2,000 scf/STB) is less than the error of 27% obtained for the whole range of GOR and 39% for high GOR wells. This high deviation is due to correlation applicability for wells with GOR less than 2,000 scf/STB.

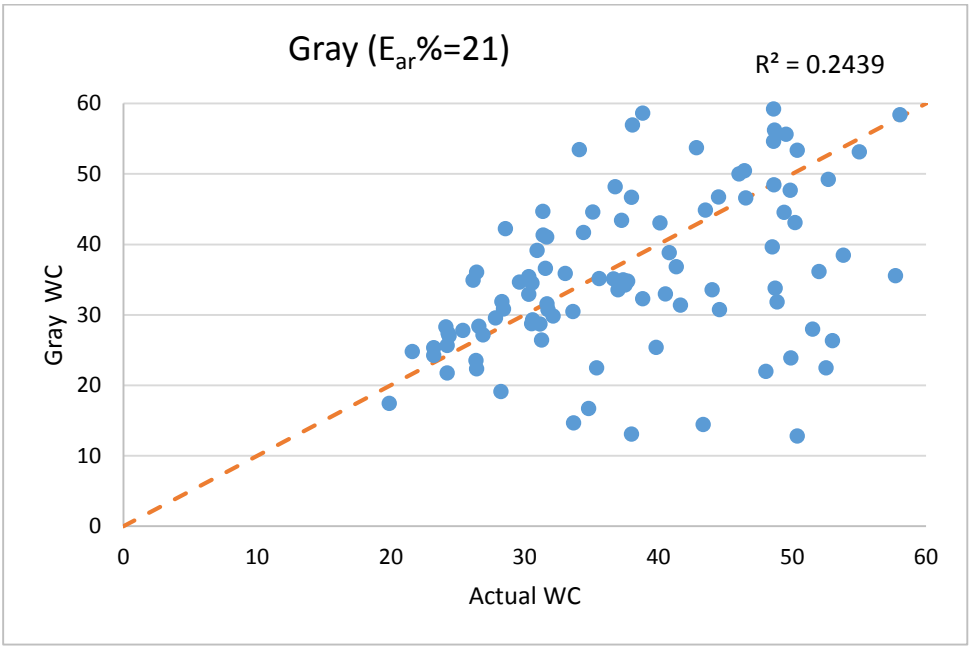


Figure 6.12 Gray WC comparison (all wells)

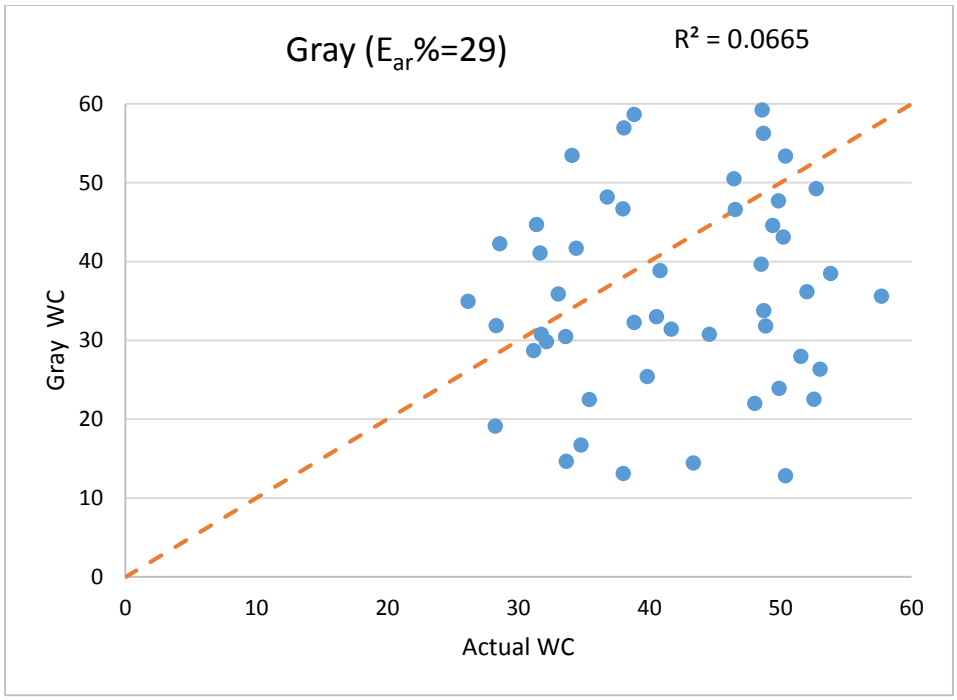


Figure 6.13 Gray WC comparison (GOR >= 2,000 scf/STB)

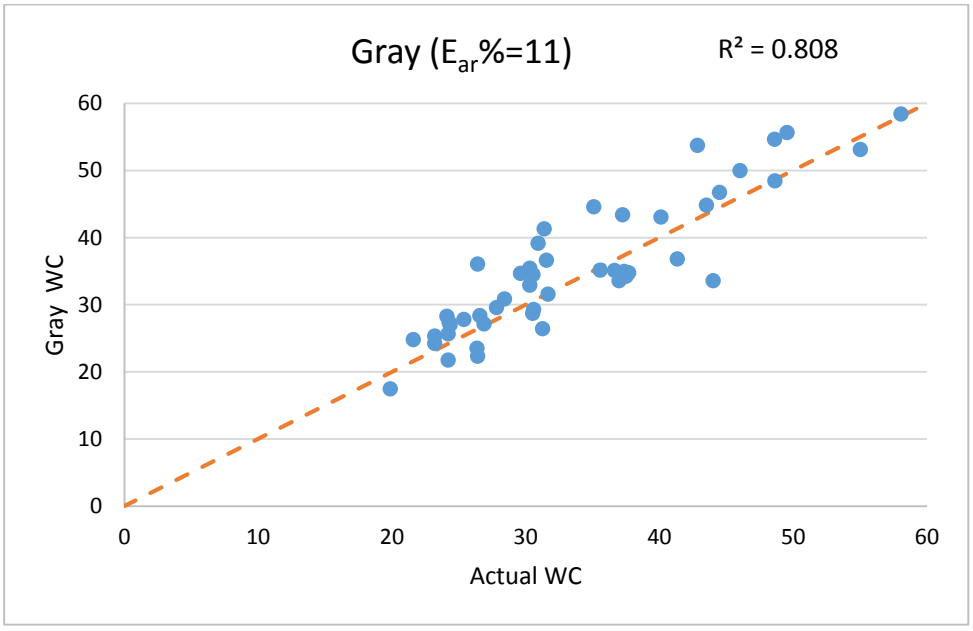


Figure 6.14 Gray WC comparison (GOR <2,000 scf/STB)

Figures 6.12, 6.13 and 6.14 exhibit the calculated water cut from Gray multiphase flow correlation against the actual water cut noting that the average absolute relative error (E_{ar}%) of 11% when applying the correlation in wells with (GOR < 2,000 scf/STB) is less than the error of 21% obtained for the whole range of GOR and 29% for high GOR wells. This deviation in error is due to correlation applicability as it gives reasonable results for wells with moderate GOR greater than 3,300 scf/STB. Gray multiphase flow correlation gives the most accurate results for black oil reservoirs among the selective five correlations in this research.

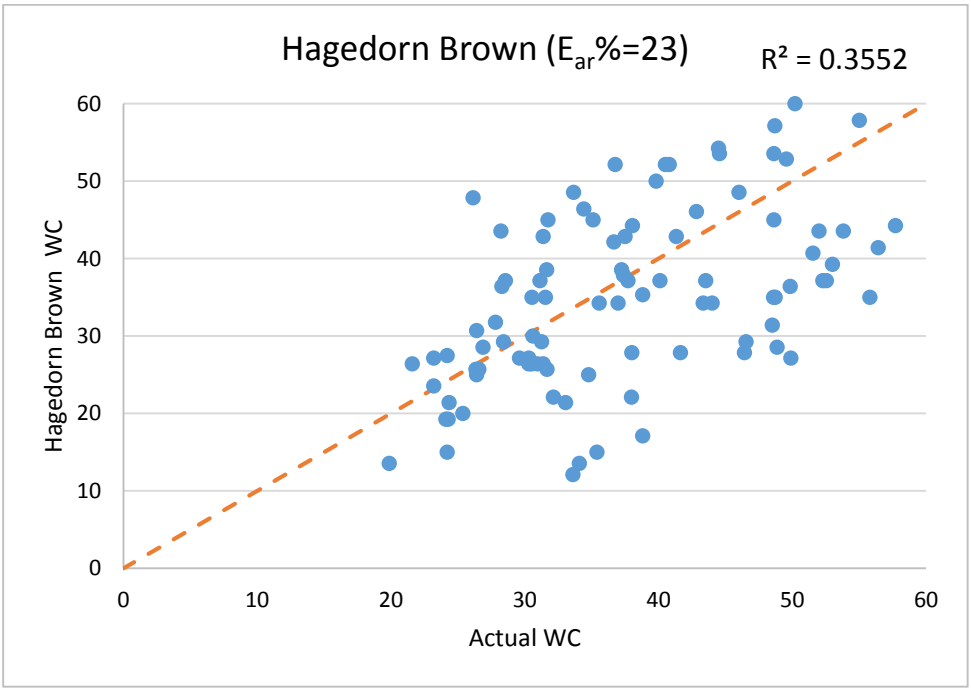


Figure 6.15 Hagedorn Brown WC comparison (all wells)

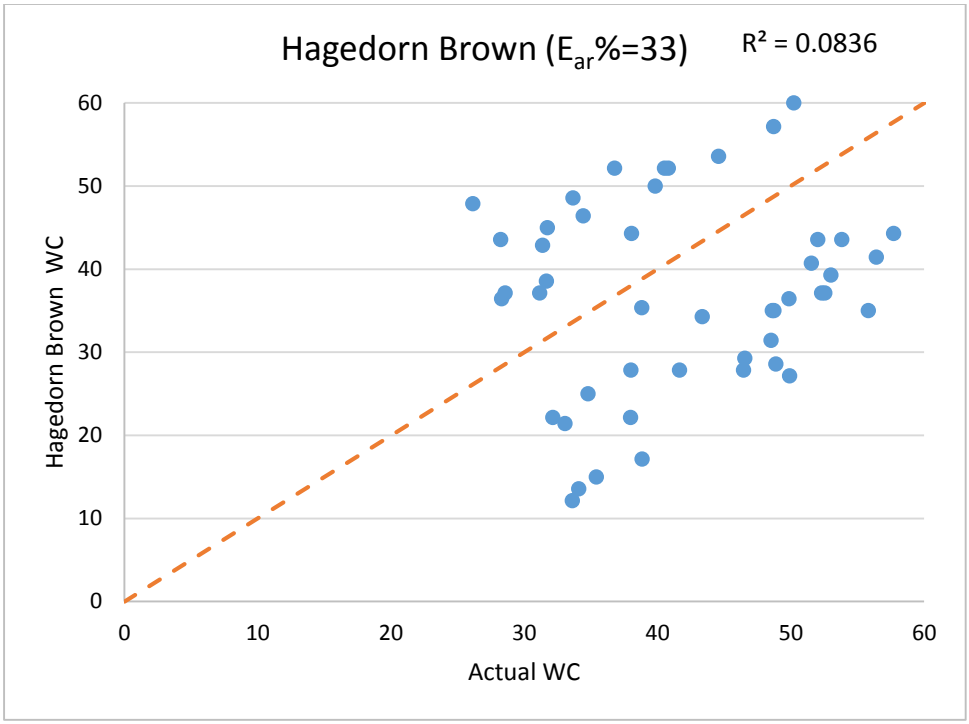


Figure 6.16 Hagedorn Brown WC comparison (all wells)

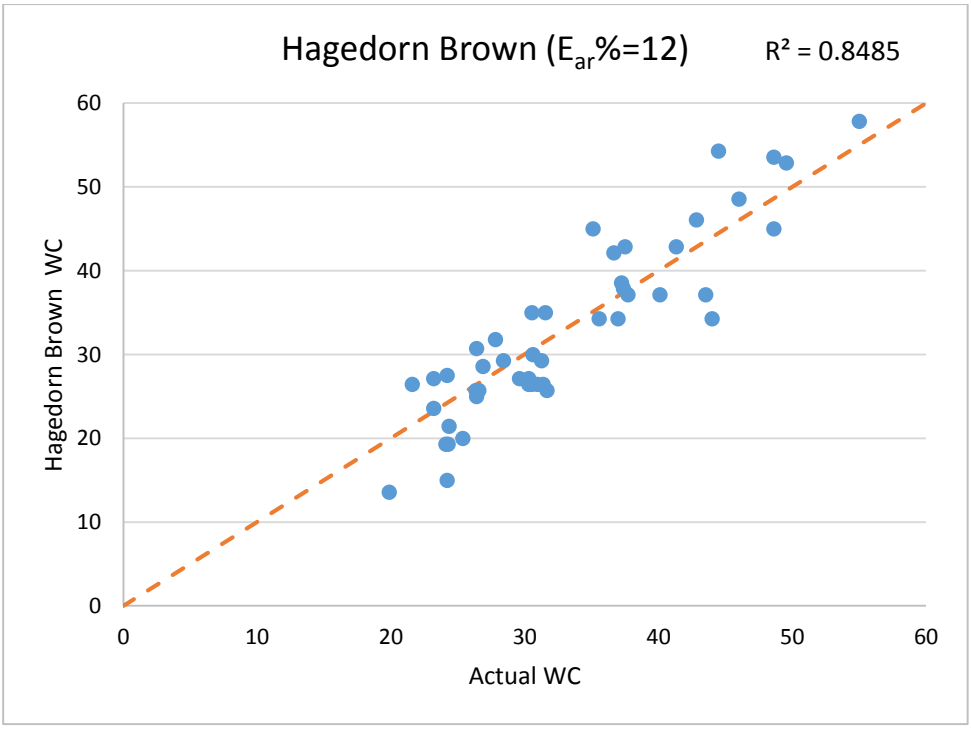


Figure 6.17 Hagedorn Brown WC comparison (GOR < 2,000 scf/STB)

Figures 6.15, 6.16 and 6.17 demonstrate the calculated water cut from Hagedorn Brown multiphase flow correlation against the actual water cut noting that the average absolute relative error (E_{ar}%) of 12% when applying the correlation in wells with (GOR < 2,000 scf/STB) is less than the error of 23% obtained for the whole range of GOR and 33% for high GOR wells. This deviation is due to correlation applicability as it gives reasonable results for wells with GOR less than 5,000 scf/STB.

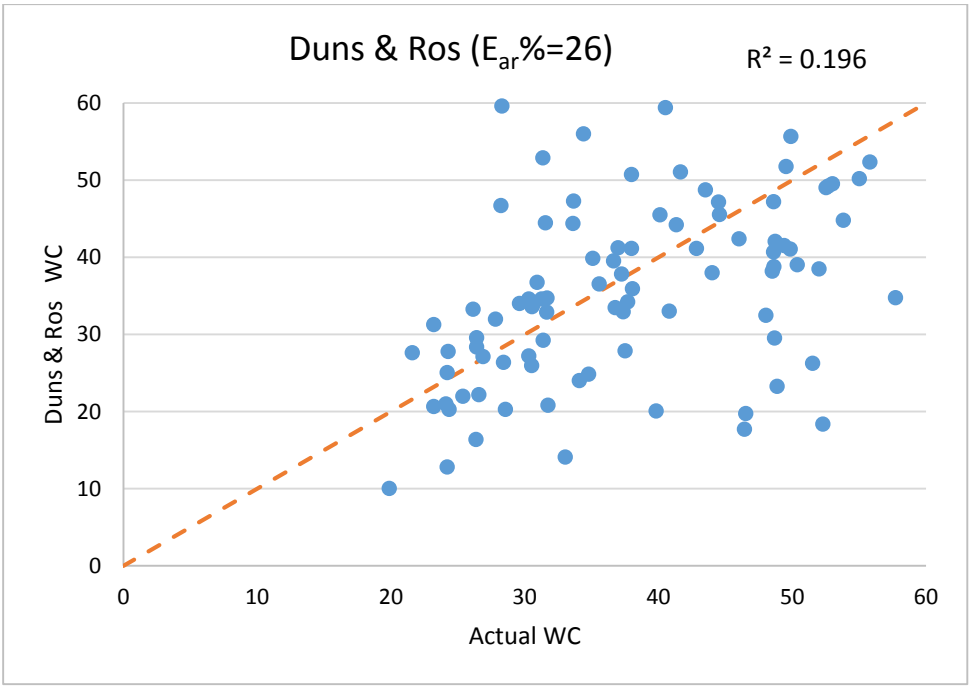


Figure 6.18 Duns & Ros WC comparison (all wells)

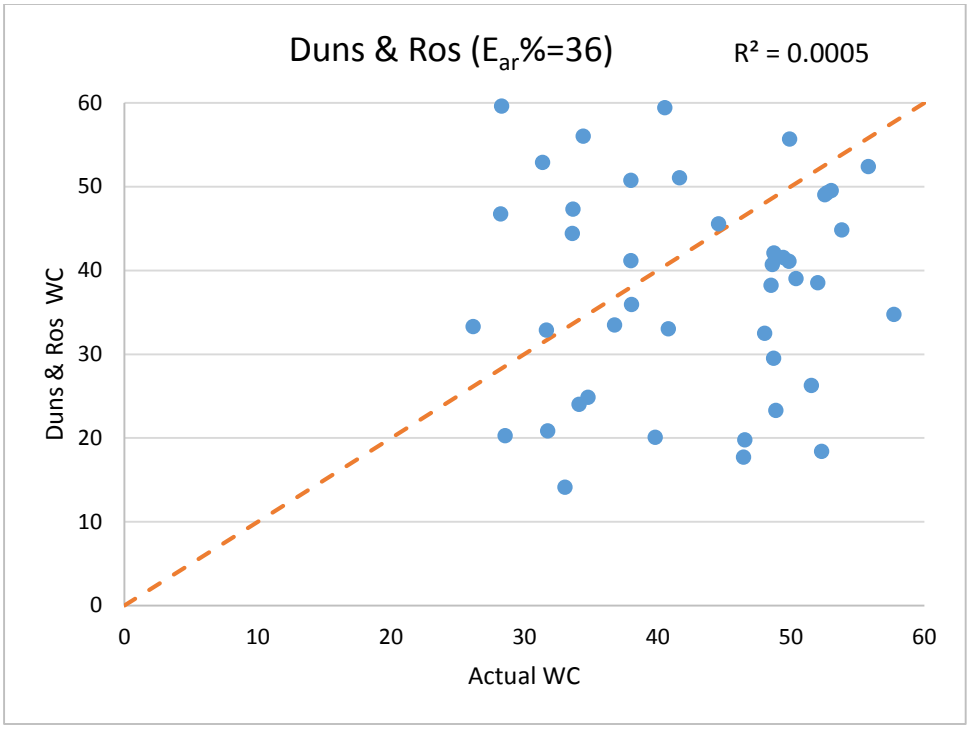


Figure 6.19 Duns & Ros WC comparison (GOR>= 2,000 scf/STB)

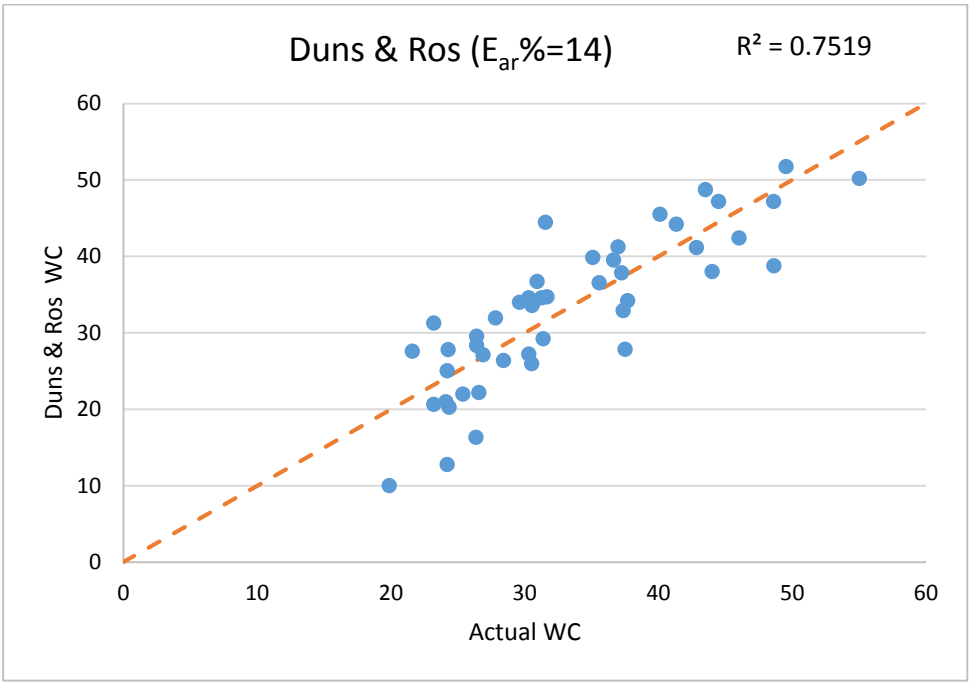


Figure 6.20 Duns & Ros WC comparison (GOR <2,000 scf/STB)

Figures 6.18, 6.19 and 6.20 show the calculated water cut from Duns & Ros multiphase flow correlation against the actual water cut noting that the average absolute relative error (E_{ar}%) of 14% when applying the correlation in wells with (GOR < 2,000 scf/STB) is less than the error of 26% obtained for the whole range of GOR and 36% for high GOR wells. The deviation in error is due to correlation applicability as it gives reasonable results for wells with GOR greater than 5,000 scf/STB but for wells with a low flow rate which is not suitable for the dataset used in this research as the flow rate span is 10,000 STB/d.

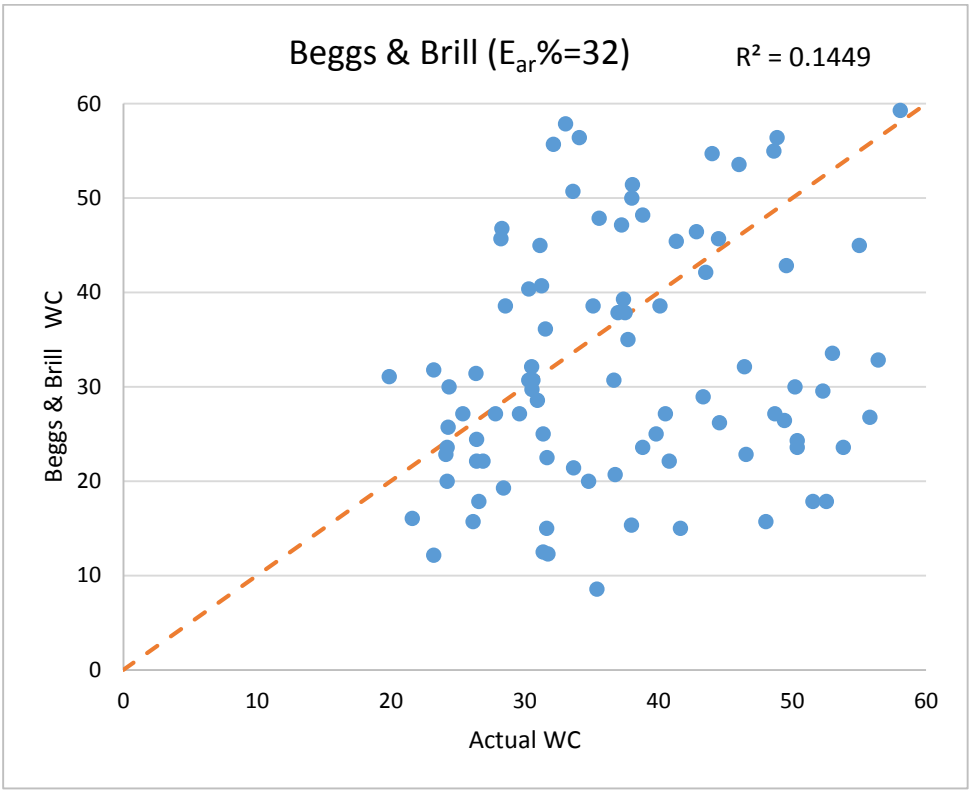


Figure 6.21 Beggs & Brill WC comparison (all wells)

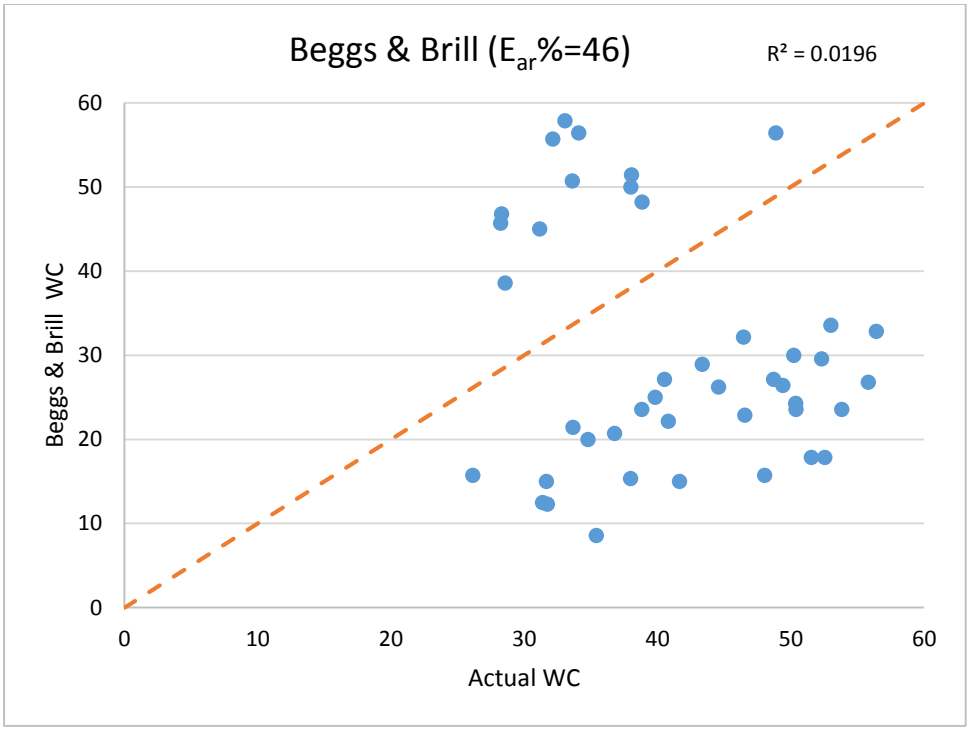


Figure 6.22 Beggs & Brill WC comparison (GOR >= 2,000 scf/STB)

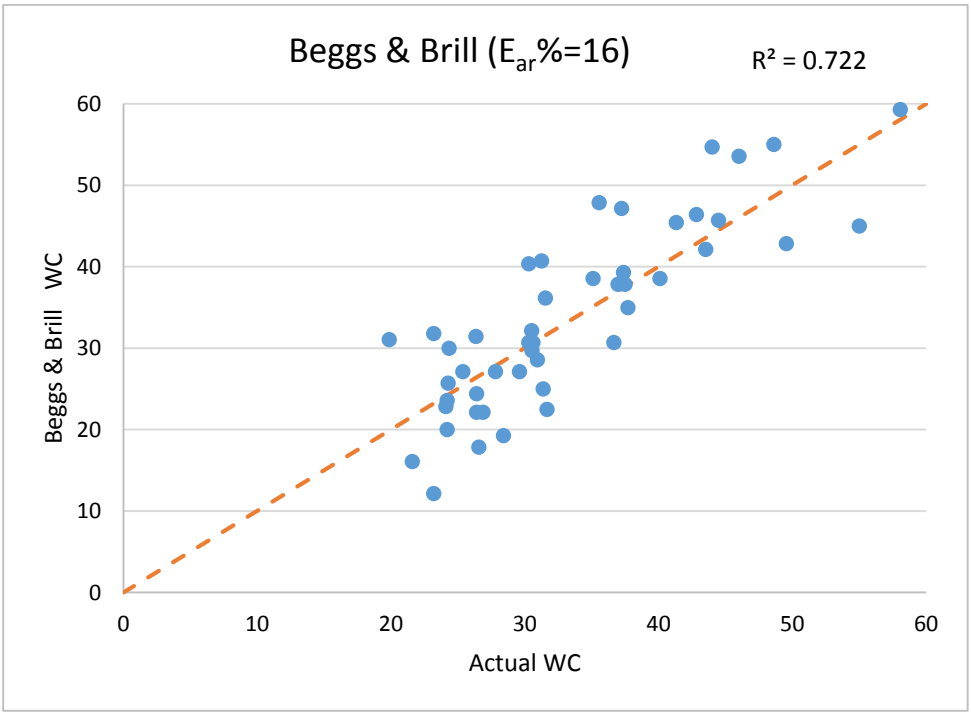


Figure 6.23 Beggs & Brill WC comparison (GOR <2,000 scf/STB)

Figures 6.21, 6.22 and 6.23 indicate the calculated water cut from Beggs & Brill multiphase flow correlation against the actual water cut noting that the average absolute relative error (E_{ar}%) of 16% when applying the correlation in wells with (GOR < 2,000 scf/STB) is less than the error of 32% obtained for the whole range of GOR and 46% error for high GOR wells (highest among the correlations). This radical error is due to correlation applicability as it overestimates the pressure drop for wells with GOR greater than 5,000 scf/STB. The subject correlation gives the highest error among the five selected correlations.

Table 6.2 summarizes the average relative percentage error for each correlation with Beggs & Brill with the highest error. It is noticeable that none of these correlations estimate water cut in high GOR wells.

Table 6.2 Multiphase flow correlations water cut errors

Multiphase flow correlation	E_{ar}% in WC		
	GOR<2,000	ALL	GOR>=2,000
Guo Ghalambor	13	27	39
Hagedorn Brown	12	23	33
Beggs & Brill	16	32	46
Gray	11	21	29
Duns & Ros	14	26	36
Average	13	25	37

A minimum error is attained by Gray correlation but it still way high. Therefore, a new empirical correlation should be developed. PROSPER™ was used to model all wells. A sample of system calculations is covered under **Appendix A**.

6.3 Models Development and Results

In this section, two items will be discussed with associated results and discussion:

- Results generated after developing new empirical correlation using nonlinear multiple regression
- Results generated after developing the ANN model

The selection criteria for the best model, with high GVF/GOR, states that the average absolute relative percentage error (E_{ar}%) should be less than 10% and average absolute percentage error (E_{aa}%) should be less than 5% as per the metering envelope in **Figure 3.5**.

6.3.1 Results from the developed empirical correlation using non-linear multiple regression

- Thirty-six (36) combinations were tested combining five inputs with below-grouped input parameters: (order of magnitude driven)

$$I_1 = \frac{P_{wh}}{P_{ds}} \quad \Big| \quad I_2 = \frac{P_{wf}}{Q_o} \quad \Big| \quad I_3 = \log(GOR)$$

- Using all available data points; the below combination was selected with minimum error attained of E_{ar}%= 20.3 and E_{aa}%= 7.7 as in **Figure 6.24**

$$\text{WC\%}=(C_1 I_1^{0.3}) + (C_2 I_2^{0.1} I_1^{0.3}) + (C_3 I_3^{0.4} I_2^{0.1}) \quad \text{..... (6.1)}$$

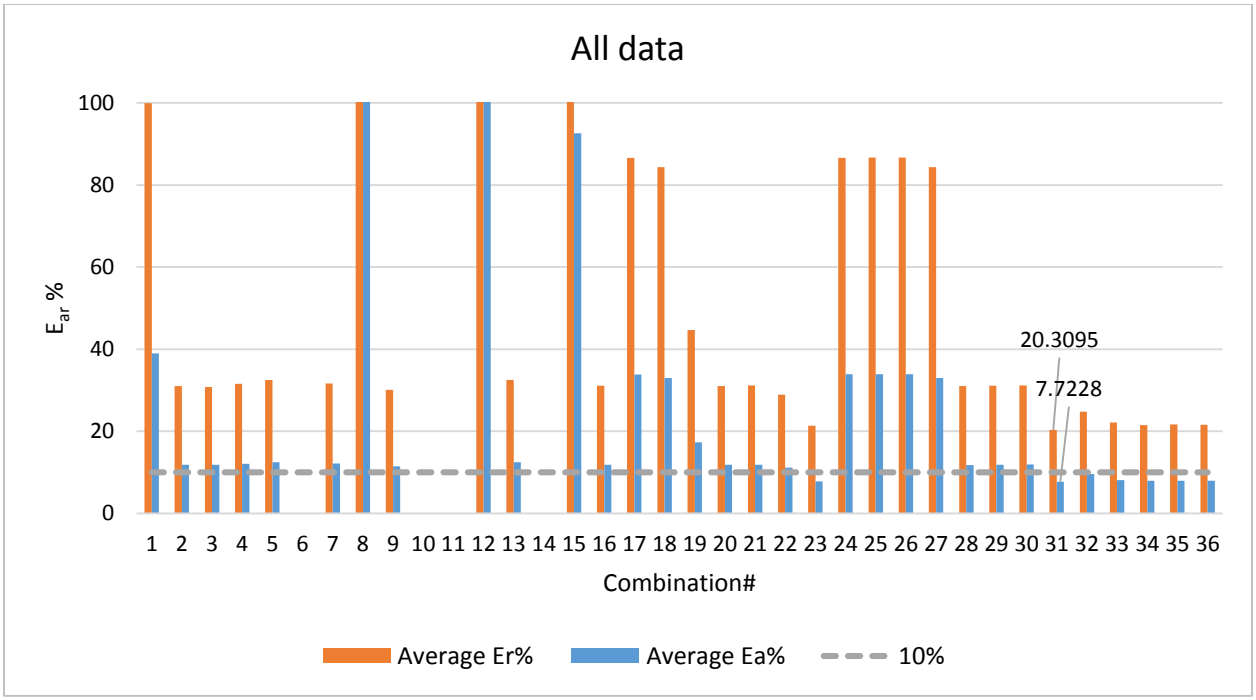


Figure 6.24 Errors for all combinations (all wells)

- Only data points with GOR <2,000 scf/STB were used with minimum error attained of E_{ar}%= 8.32 (<10%) & E_{aa}%= 2.96 (<5%)

$$WC\%=(-134.7881 I_1^{0.3}) + (195.1021 I_2^{0.1} I_1^{0.3}) + (-6.2683 I_3^{0.4} I_2^{0.1}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6.2)$$

The developed correlation is only working for wells with GOR less than 2,000 scf/STB. Therefore, another method needs to be in place to accommodate the whole range. **Figure 6.25** illustrates errors obtained from all combinations. Also, data with GOR greater than or equal 2,000 scf/STB were tested. The E_{ar} was 25.1%, and E_{aa} was 9.4% which is relatively very high.

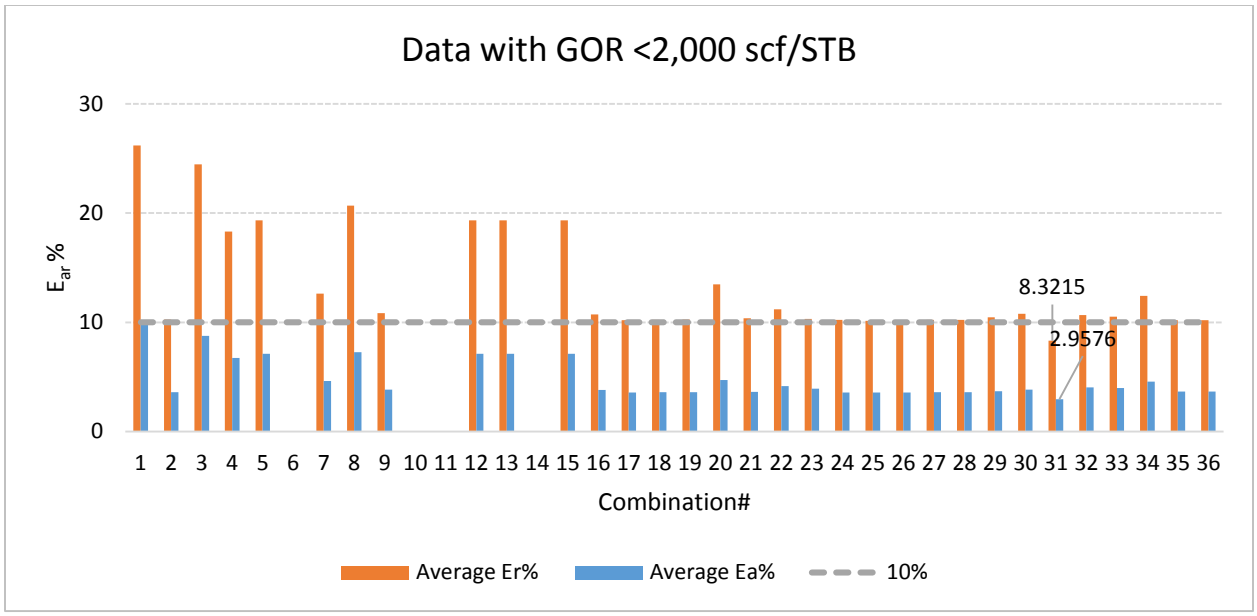


Figure 6.25 Errors for all combinations (GOR < 2,000 scf/STB)

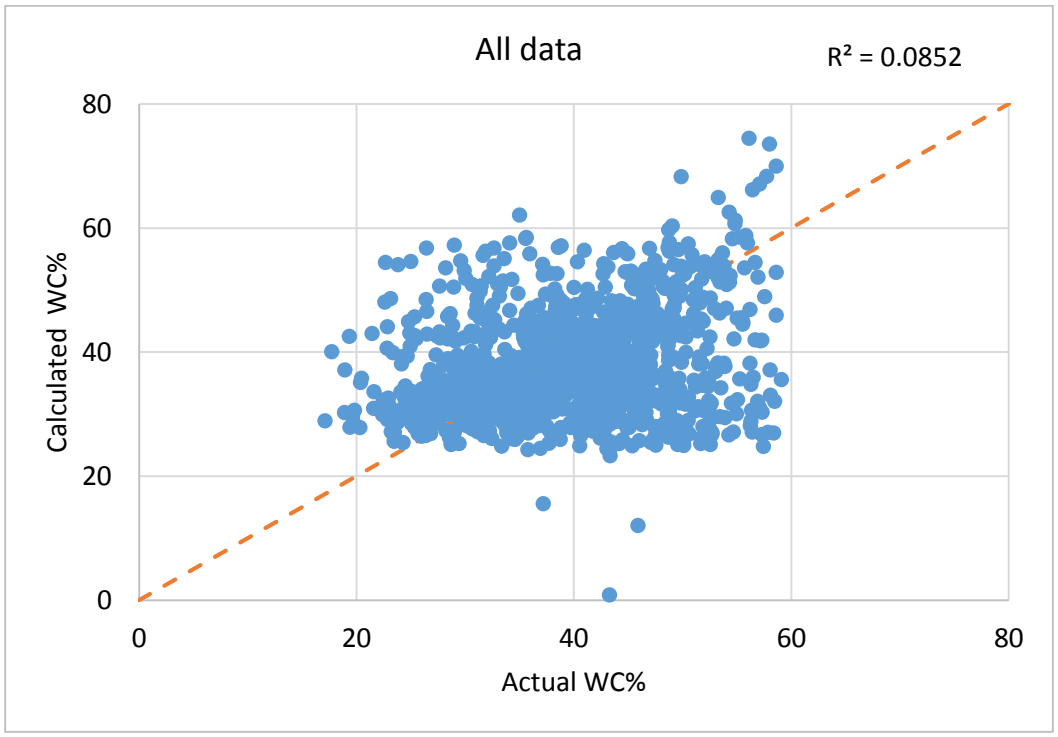


Figure 6.26 cross-plot of actual and calculated water cut (all wells)

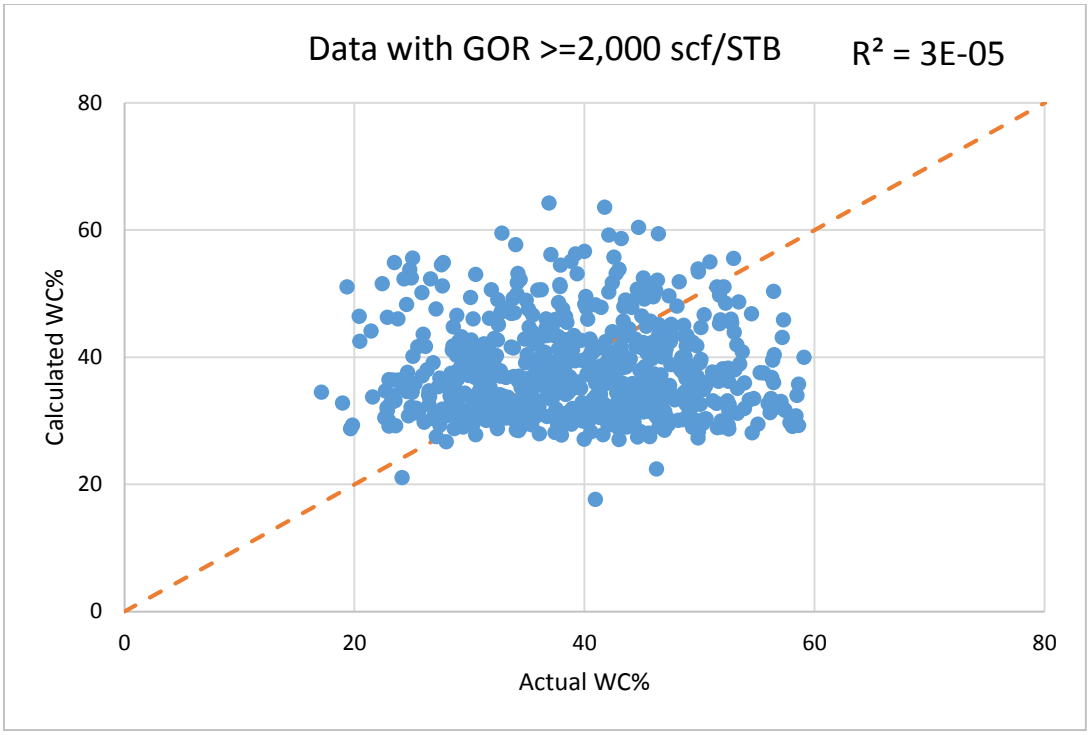


Figure 6.27 cross-plot of actual and calculated water cut (wells with GOR >=2,000 scf/STB)

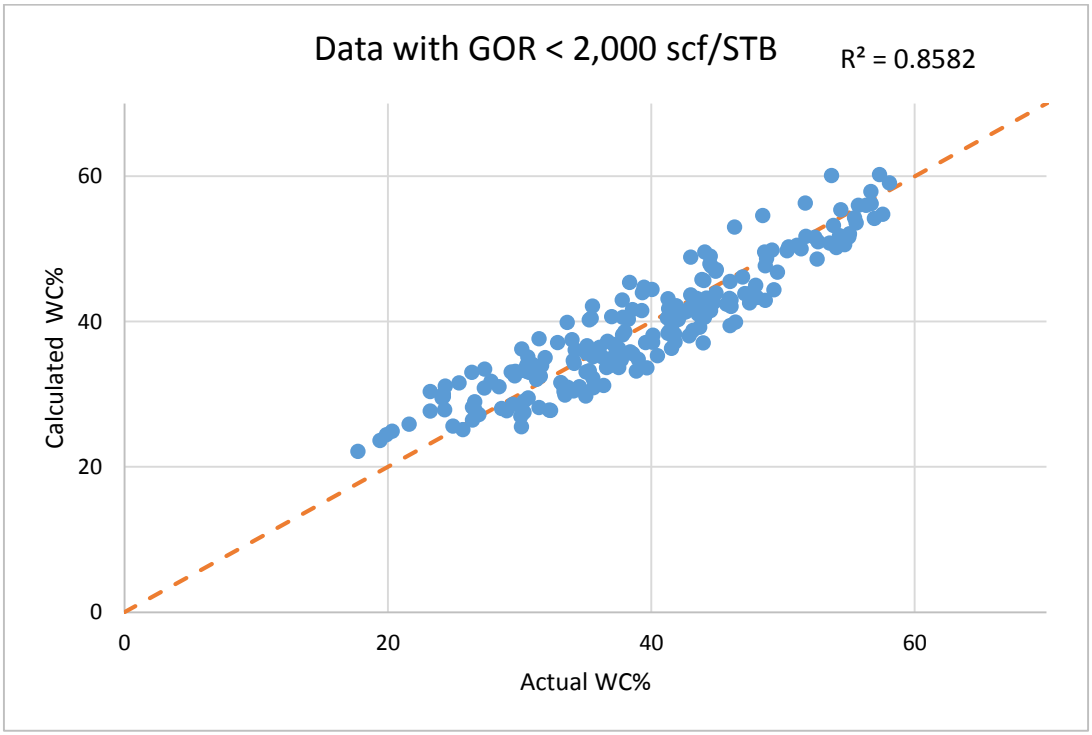


Figure 6.28 cross-plot of actual and calculated water cut (wells with GOR <2,000 scf/STB)

Figures 6.26 and 6.27 show the high deviation and error with using the new developed empirical correlation for all wells and wells with GOR of greater than 2,000 scf/STB. Both cross-plots exhibit E_{ar}% above 20%.

Figure 6.28 demonstrate the achieved accuracy and lower error from using the new developed empirical correlation for wells with GOR of less than 2,000 scf/STB.

Table 6.3 Empirical Correlation Development Attempts with Error Statistics

Combination#	ALL		<2,000 scf/STB	
	Average E_{ar}%	Average E_{aa}%	Average E_{ar}%	Average E_{aa}%
1	99.9182	38.9964	26.2028	9.8342
2	31.0392	11.8039	10.2568	3.6079
3	30.7583	11.8361	24.4717	8.7595
4	31.5466	12.0626	18.3276	6.7355
5	32.5018	12.438	19.3269	7.1214
6				
7	31.6128	12.1629	12.6223	4.618
8	5.17E+13	2.83E+13	20.6979	7.2608
9	30.0616	11.4307	10.8426	3.8475
10				
11				
12	3.13E+08	9.48E+07	19.3269	7.1214
13	32.5018	12.438	19.3269	7.1214
14				
15	282.6054	92.5948	19.3269	7.1214
16	31.1167	11.8495	10.7378	3.817
17	86.5936	33.8407	10.1858	3.58
18	84.3081	32.9515	10.2267	3.6009
19	44.666	17.2621	10.246	3.6059
20	31.0439	11.8057	13.4767	4.7123

Combination#	ALL		<2,000 scf/STB	
	Average E_{ar}%	Average E_{aa}%	Average E_{ar}%	Average E_{aa}%
21	31.1843	11.8685	10.361	3.6448
22	28.9138	11.0997	11.1862	4.1594
23	21.3379	7.7837	10.3228	3.9311
24	86.6187	33.8607	10.2181	3.5905
25	86.6814	33.8977	10.1513	3.5695
26	86.706	33.8901	10.1329	3.5637
27	84.3081	32.9515	10.2267	3.6009
28	30.9805	11.7883	10.2267	3.5993
29	31.0597	11.8131	10.4518	3.689
30	31.2056	11.8766	10.7878	3.8423
31	20.3095	7.7228	8.3215	2.9576
32	24.8116	9.5978	10.6625	4.05
33	22.1107	8.0776	10.5064	3.9965
34	21.4881	7.898	12.4237	4.5861
35	21.6171	7.9402	10.2059	3.6603
36	21.5627	7.9073	10.2068	3.6625

Table 6.3 demonstrates the average absolute error and average absolute relative percentage error of each empirical correlation combination considering two scenarios; all wells and wells with low GOR. As shown in **Figure 6.27**, wells with high GOR exhibit high deviation and, therefore, combinations were not tested for this scenario. It is worth noting that combinations #6, 10, 11 and 14 failed to work and generate results and so they cannot be applied in this research. Nevertheless, combination #31 shows the lowest error where it leaves it as the choice for the correlation. It is only applicable for black oil reservoirs where GOR is less than 2,000 scf/STB. Now, there is a need to find another technique to estimate water cut covering all range of GOR. In this case, ANN will be an alternative technique.

6.3.2 Results from the developed Artificial Neural Network Model

I. Initial Run

- The selected five inputs from random forest results used in this model with no parameter grouping.
- Three layers in the model as in **Figure 6.29**:
  - i. Input layer which has nine features/inputs
  - ii. Hidden layer; 1 hidden layer with ten neurons
  - iii. Output layer has one output with one neuron

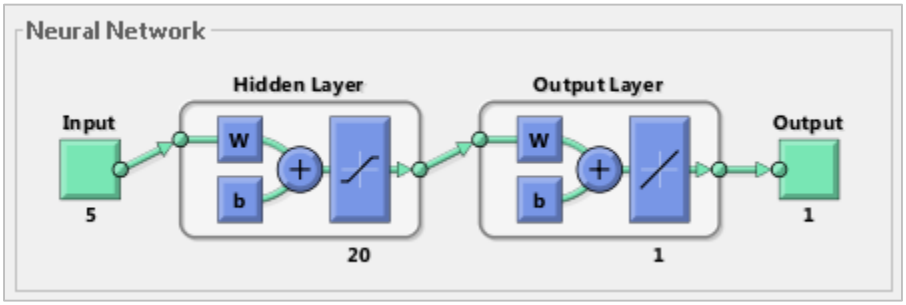


Figure 6.29 NN Structure (five inputs)

- Hidden layer transfer (activation) to activate neurons; tangent sigmoid is used because it is best to predict numerical output and it is the default transfer function.
- Output layer transfer function is linear since there is only one neuron.
- The feed-forward back-propagation algorithm was used to carry out NN.

- The training/testing function used is Levenberg Marquardt which is considered to be the fastest algorithm.
- One thousand two hundred ten (1,210) data points were used, out of which 70% was used for training and 30% was used for testing.
- The E_{ar}% attained from the initial run was 8.42%, and the average E_{aa}% was 3.03%.

II. Training Process Optimization Results

The optimization process is subdivided into three steps:

i. Optimizing different training functions

Multiple training functions were tried out with fixing the number of neurons of 10 in the hidden layer (s) and using hyperbolic tangent sigmoid transfer function as detailed in section 4.2.

Appendix B.I covers **Figures B.1 to B.9** which illustrate the cross-plot of water cut calculated and measured for each training function.

Table 6.4 delineates the best training function to be utilized in ANN model where the Levenberg Marquardt with one hidden layer is the selected training function as it has a low E_{ar}% where the error obtained from training got reduced in the testing which is a good indicator of how good the training function is.

Table 6.4 Summary table for training functions' optimization

Type#	E_{ar}%_tr	E_{ar}%_ts	Avg E_{ar}%	Avg E_{aa}%	R^2
0	8.48	8.06	8.4169	3.0331	0.8352
1	8.161	8.908	8.2733	2.9891	0.8403
2	130.489	137.539	131.5495	45.3850	0.0838
3	9.549	9.629	9.5613	3.4934	0.7692

Type#	E_{ar}%_tr	E_{ar}%_ts	Avg E_{ar}%	Avg E_{aa}%	R^2
4	8.308	8.405	8.3227	3.0054	0.838
5	8.547	9.027	8.6195	3.1099	0.8207
6	8.714	8.616	8.6988	3.1380	0.819
7	8.266	8.626	8.3203	3.0021	0.8360
8	8.399	8.409	8.4006	3.0322	0.8323

ii. Optimizing different transfer functions

Multiple transfer functions were tried out with fixing the number of neurons of 10 in the hidden layer and using a type 0 training function. Transfer functions are hyperbolic tangent sigmoid (tansig), log sigmoid (logsig), pure linear (purelin), symmetric hard limit (hardlims), and symmetric saturated linear (satlins) with Appendix B.II **Figures B.10 to B.14.**

Table 6.5 shows that symmetric saturated linear transfer function performs best for the subject ANN model while hard limit function performs very poorly.

Table 6.5 Statistical error for transfer functions' optimization

Transfer function name	E_{ar}%_tr	E_{ar}%_ts	Avg E_{ar}%	Avg E_{aa}%	R^2
tansig	8.48	8.06	8.4169	3.0331	0.8352
logsig	8.527	8.19	8.4761	3.0875	0.8355
purelin	9.438	8.883	9.3546	3.3469	0.7871
hardlims	15.774	15.778	15.7745	5.6955	0.3675
satlins	8.356	8.071	8.3133	3.0119	0.8367

iii. Optimizing different neurons count in the hidden layer

A hidden layer neurons count from 1 to 25 was attempted with the objective of getting the lowest error with fixing the training and transfer function. As per **Figure 6.30 and Table 6.6**, neurons of 20 are the most optimized count since it has the lowest E_{ar}%.

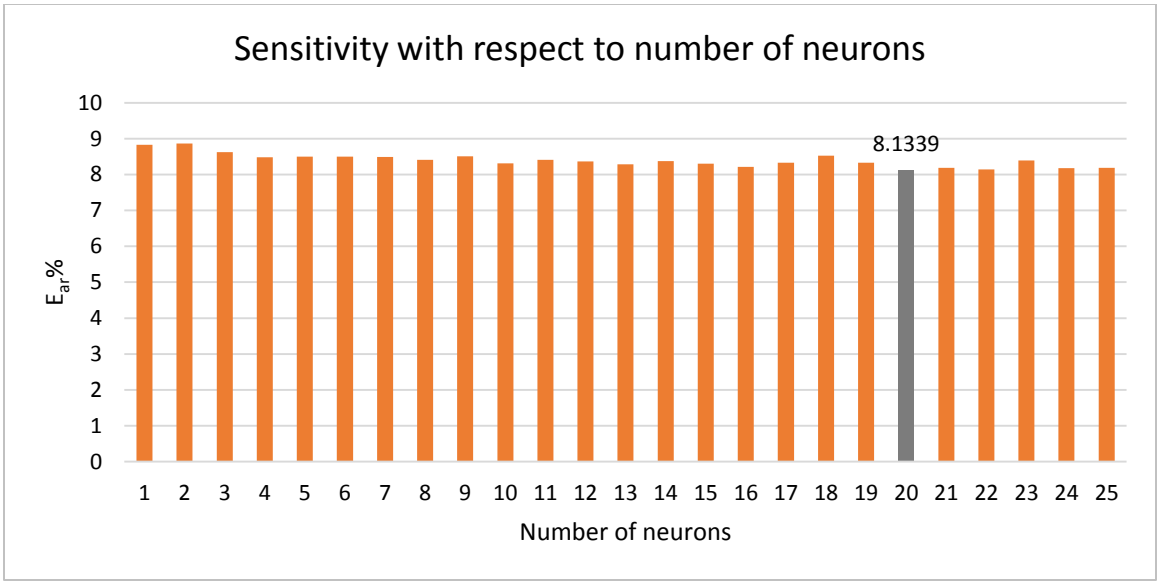


Figure 6.30 Sensitivity analysis with changing neurons count in the hidden layer

Table 6.6 Statistical error for neuron counts' optimization in the hidden layer

No. of neurons	Avg E_{ar}%	Avg E_{aa}%	R^2
1	8.8332	3.1765	0.8155
2	8.8661	3.1710	0.8142
3	8.6222	3.0905	0.828
4	8.4831	3.0575	0.8312
5	8.4976	3.0408	0.8298
6	8.4981	3.0379	0.8334
7	8.4897	3.0522	0.8318
8	8.4134	3.0556	0.8352
9	8.5117	3.1062	0.8293
10	8.3133	3.0119	0.8367

No. of neurons	Avg E_{ar}%	Avg E_{aa}%	R^2
11	8.4149	3.0451	0.8335
12	8.3662	3.0137	0.8346
13	8.2869	2.9973	0.8367
14	8.3752	2.9993	0.8314
15	8.3086	2.9843	0.8368
16	8.2190	2.9854	0.8388
17	8.3345	2.9916	0.8341
18	8.5270	3.0691	0.8294
19	8.3323	3.0039	0.8357
20	8.1339	2.9315	0.8414
21	8.1847	2.9389	0.8388
22	8.1482	2.9470	0.8412
23	8.3941	3.0223	0.8338
24	8.1824	2.9746	0.8398
25	8.1846	2.9515	0.8385

III. Optimization Implementation and Results

After completing the optimization process, the optimized ANN model was used with five inputs and another with nine inputs to observe the difference between the two cases:

i. ANN model with five inputs

The optimized ANN model is used with five inputs/neurons in the input layer, 20 neurons in the hidden layer, and one neuron in the output layer. Symmetric saturating linear transfer and Levenberg–Marquardt training functions were used as per **Table 6.7**.

Table 6.7 Final ANN Model Characteristics with Five Inputs for all Wells

FINAL OPTIMIZED ANN MODEL	
Input Parameters	5
Output Parameter	1
Training Function	Levenberg-Marquardt (trainlm)
Number of Hidden Layers	1
Hidden Layer Transfer Function	Symmetric Saturating Linear (satlins)
Number of Neurons in Input Layer	5
Number of Neurons in Hidden Layer	20
Number of Neurons in Output Layer	1

ANN empirical equation with five inputs:

$$(Water\ Cut)_n = [\sum_{i=1}^{N} w_{2i} * satlins(w_{1i,j}x_j + b_{1i})] + b_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6.3)$$

Where i denoted for neuron trace and j for input trace

Where x is the normalized input (P_{wh}, P_{ds}, P_{wf}, GOR & Q_o)

Figures 6.31 to 6.33 capture the trend of training, testing and overall dataset for the ANN model with five inputs for all wells.

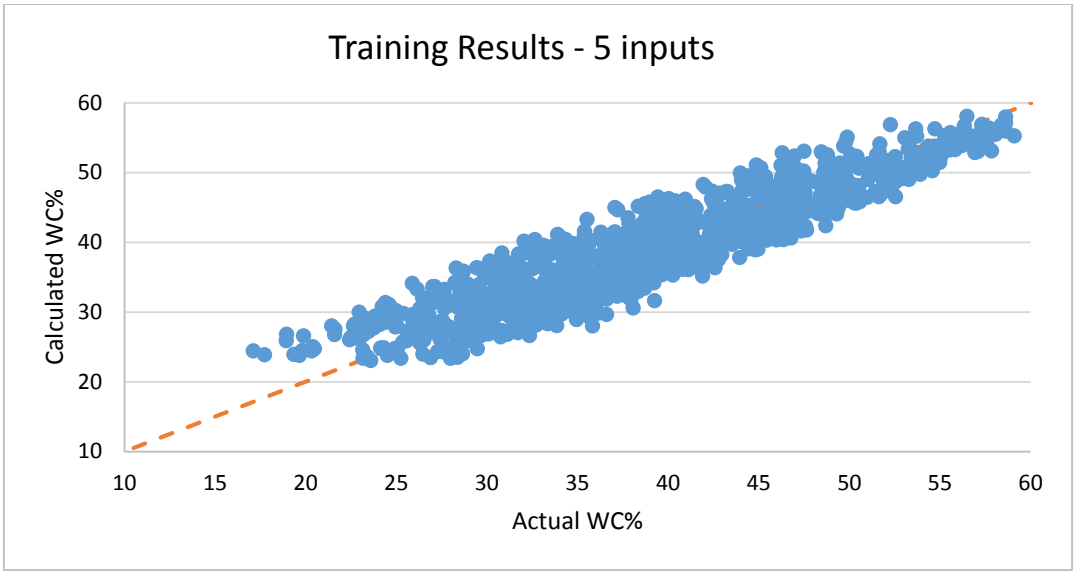


Figure 6.31 Training results with the ANN model of five inputs for all wells

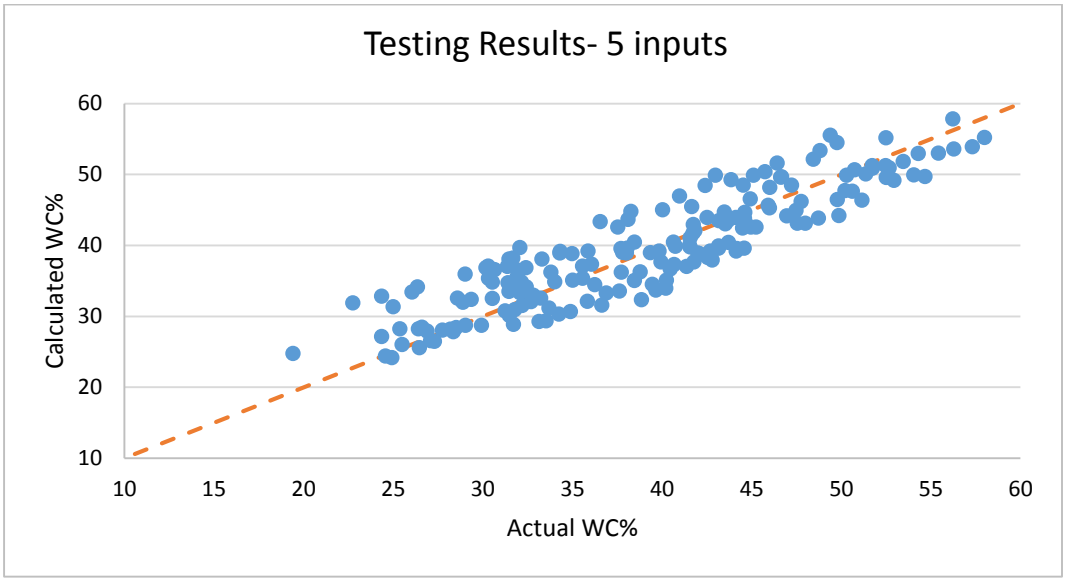


Figure 6.32 Testing results with the ANN model of five inputs for all wells

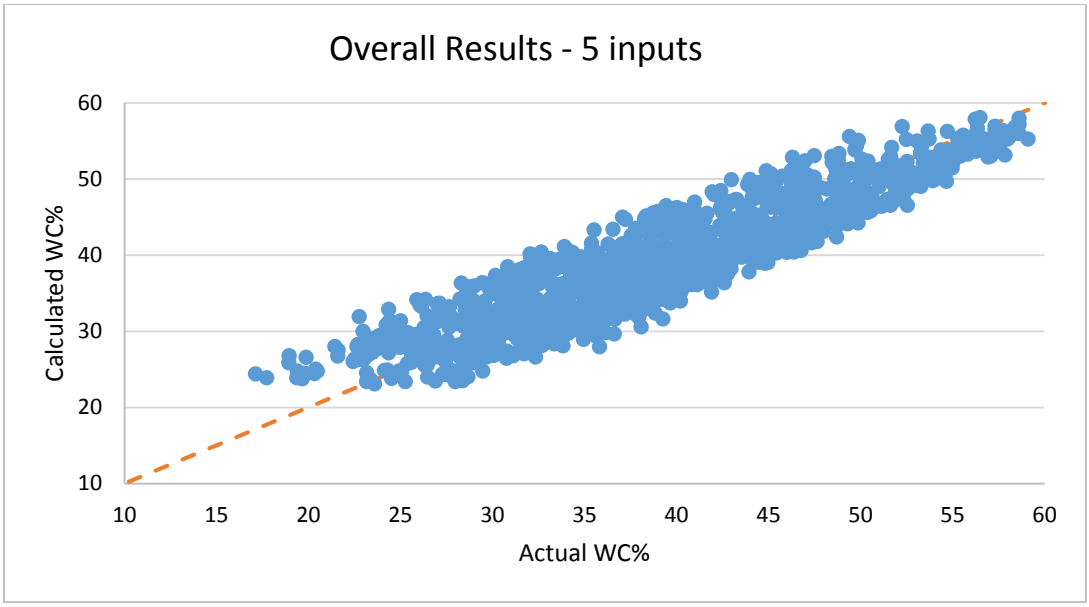


Figure 6.33 Overall results with the ANN model of five inputs for all wells

Table 6.8 reflects the error analysis resulted from running the ANN model with five inputs. Overall error E_{ar} is 8.13% which is less than the acceptable error of 10%. Also, the absolute error E_{aa} is less than 5% which demonstrates the accuracy of the developed model with only five inputs.

Table 6.8 Error Analysis with Five Inputs for all Wells

Stage	**E_{ar}%**	**E_{aa}%**	**CC**	**R^2**
Training	8.1129	2.9153	0.92	0.85
Testing	8.2523	3.0232	0.90	0.81
Overall	8.1339	2.9315	0.92	0.84

Additionally, **Table 6.9** reveals all weights and biases with the normalized form of the empirical equation to the target.

Table 6.9 Weights and biases for the final form (five inputs)

Number of Neurons in the hidden layer	Weights (w_1)					Weight (w_2)	Bias (b_1)	Bias (b_2)
	P_{wh}	P_{ds}	Q_o	GOR	P_{wf}	WC%	ALL inputs	WC%
1	-0.8803	1.8154	3.042	0.6576	-1.7758	-0.028	-0.6028	0.6422
2	2.6361	-3.6259	3.9233	-3.0761	2.039	-0.0661	-4.2973	
3	1.3058	0.3943	-1.2482	-1.4707	-0.0973	0.0981	-2.0378	
4	0.2695	-0.3222	0.9158	-0.0869	1.1345	-0.1391	-5.0317	
5	-0.3919	0.5023	-2.6813	-1.0285	-1.5317	0.0468	0.3559	
6	0.4499	-2.1958	1.4376	0.4723	-1.5815	-0.0177	0.411	
7	-0.2977	0.2497	1.757	0.0464	0.3204	-0.5482	2.0617	
8	0.8897	-2.1076	-4.7889	-1.1335	-0.8897	0.006	-0.9911	
9	-0.8529	1.3077	-1.155	0.4627	-0.0885	-0.0147	-0.9634	
10	-1.7907	2.1857	-1.9953	1.0042	0.185	0.1062	-0.3542	
11	1.3267	1.0078	3.636	0.3614	-1.6818	0.1274	2.2107	
12	-0.9402	-2.4072	-0.9218	3.8994	-2.243	-0.0621	-3.1724	
13	0.6548	-0.1091	-3.3397	0.7967	0.0541	-0.0393	-1.3422	
14	-1.4405	3.3782	0.9861	-0.7787	0.0693	-0.0781	-2.077	
15	-0.9019	-0.5357	-2.0223	0.0246	1.1546	0.5736	-1.3515	
16	0.8363	0.93	-1.1114	-0.2312	0.7943	-0.1101	-2.8228	
17	-0.7959	1.9978	0.407	-0.1513	0.7789	0.012	-1.3888	
18	-0.6634	1.7659	-0.3781	4.0804	-0.1645	0.0091	-2.0969	
19	-0.417	1.0138	-0.4872	0.1957	-0.1752	0.3881	-1.4146	
20	1.3454	-1.4729	-3.4763	-0.0988	-0.4377	0.1434	0.856	

w_1: Weights vector between input and the hidden layer of the neural network.

& w_2: Weights vector between the hidden and the output layer of the neural network.

b_1: Bias vector between input and the hidden layer of the neural network.

& b_2: Bias between the hidden and the output layer of the neural network.

The normalized empirical equation for the ANN model with five inputs

$$(\text{Water Cut})_{calc} = 0.84126 * (\text{Water Cut})_{actual} + 6.2528 \quad \ldots\ldots (6.4)$$

Figures 6.34 and 6.35 represent the overall results for low GOR and high GOR wells, respectively. The Average Absolute Relative Percentage Error for low GOR wells is 7.2% while the error for high GOR wells is 8.4%. Both give an error less than the acceptable limit of 10%, which confirm the ANN model accuracy with five inputs.

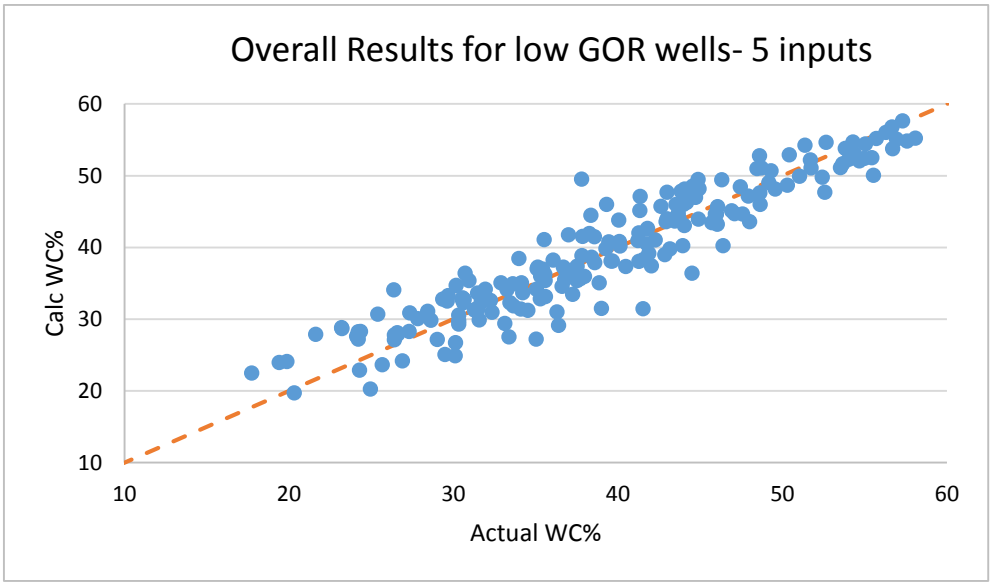


Figure 6.34 Overall results with the ANN model of five inputs for low GOR wells

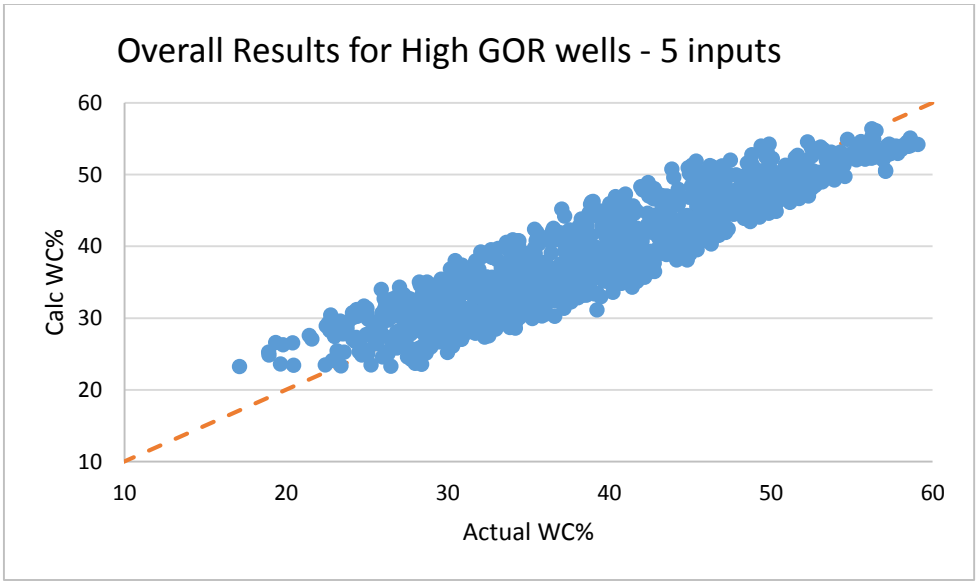


Figure 6.35 Overall results with the ANN model of five inputs for high GOR wells

ii. ANN model with nine inputs

The optimized ANN model is used with nine inputs/neurons in the input layer, ten neurons in the hidden layer and one neuron in the output layer. Hyperbolic tangent sigmoid transfer and Levenberg–Marquardt training functions were used as per **Table 6.10**.

Table 6.10 Final Optimized ANN Model Characteristics with Nine Inputs

Input Parameters	9
Output Parameter	1
Training function	Levenberg-Marquardt (trainlm)
Number of hidden layers	1
Hidden layer Transfer function	Hyperbolic Tangent Sigmoid (tansig)
Number of neurons in the input layer	9
Number of neurons in the hidden layer	10
Number of neurons in the output layer	1

ANN empirical equation with nine inputs:

$$(Water\ Cut)_n = [\sum_{i=1}^{N} w_{2i} * tansig(w_{1i,j}x_j + b_{1i})] + b_2 \dots\dots\dots\dots\dots\dots\dots\dots (6.5)$$

Where x is the normalized input parameter, i denoted for neuron trace and j for input trace

Figures 6.36 to 6.38 capture the trend of training, testing and overall dataset for the ANN model with nine inputs.

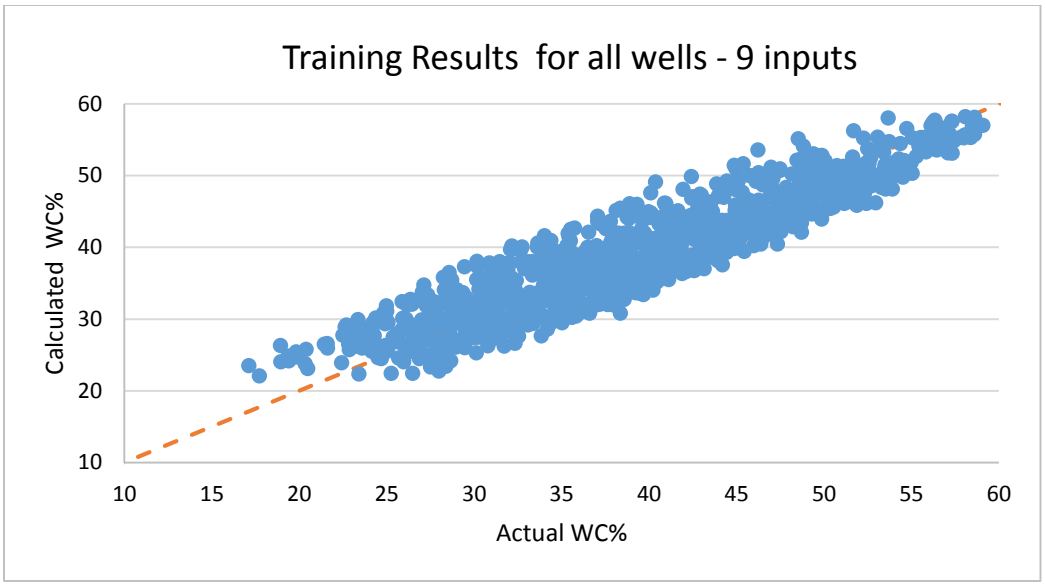


Figure 6.36 Training results with the ANN model of nine inputs for all wells

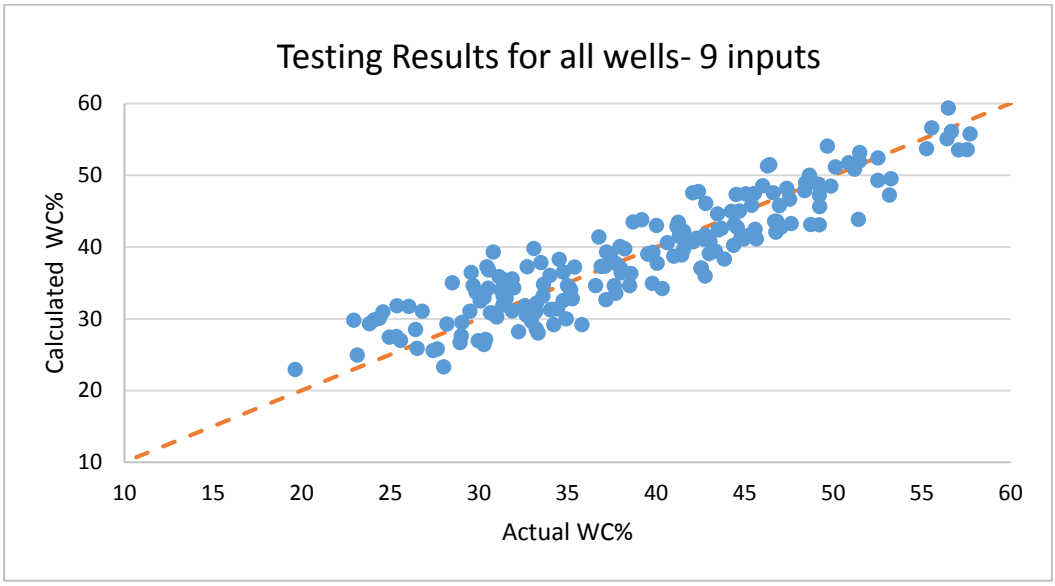


Figure 6.37 Testing results with the ANN model of nine inputs for all wells

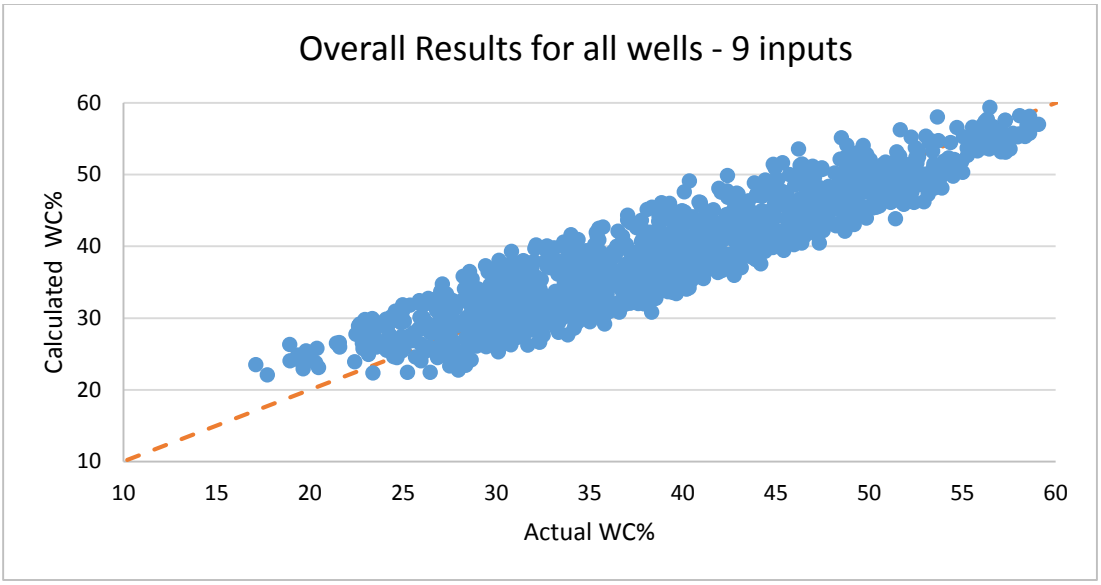


Figure 6.38 Overall results with the ANN model of nine inputs for all wells

Table 6.11 Error Analysis with Nine Inputs for all Wells

Stage	E_{ar}%	E_{aa}%	CC	R^2
Training	8.2375	2.9715	0.93	0.86
Testing	7.8787	2.8201	0.91	0.84
Overall	8.1835	2.9487	0.92	0.85

Table 6.11 reflects the error analysis resulted from running the ANN model with nine inputs. The overall error E_{ar} is 8.18%, which is less than the acceptable error of 10%. Also, the absolute error E_{aa} is less than 5% which demonstrates the accuracy of the developed model with only nine inputs.

It is evident that the ANN model with nine inputs resulted with average E_{ar}% of 8.18 against an error attained of 8.13 when using ANN model with five inputs. This difference in error makes no significant difference when using either five or nine inputs as the selected five inputs are the most

important variables from the Random Forest technique. Weights and biases are tabulated in **Table 6.12** with the normalized empirical equation to the target.

Table 6.12 Weights and biases for the final form (nine inputs)

Number of Neurons in the hidden layer	Weights (w1)									Weight (w2)	Bias (b1)	Bias (b2)
	Choke%	P_{wh}	P_{ds}	T_{wh}	Q_o	GOR	P_{wf}	T_{wf}	ICV Position	WC%	All inputs	WC%
1	1.3362	-1.058	0.1675	1.1744	-0.6551	-1.2278	0.3698	0.3058	-0.4694	0.0469	-2.6369	1.1344
2	2.8891	1.1151	4.9803	-2.4957	-0.286	-1.1838	0.0026	-3.5722	-2.7394	0.0521	-4.3794	
3	-0.4158	-0.9538	1.3859	-0.2235	1.2521	0.019	-1.0769	-1.2409	-0.5719	-0.0105	1.7605	
4	0.2114	-0.2288	0.7477	-0.5191	1.7661	0.3076	-0.7346	0.8042	-1.33	0.3181	-0.1905	
5	-0.3795	1.2896	-0.5095	-0.6378	-0.3897	0.3935	0.1461	-0.8536	-1.3111	0.023	1.0748	
6	1.0919	-0.2626	0.2707	0.2146	-0.2846	1.689	0.3106	0.3569	0.0665	0.0721	0.0646	
7	0.0947	0.0813	0.0759	-0.6812	0.8333	0.1323	-0.1404	0.7166	-0.0684	-1.6231	0.8509	
8	-1.1472	-1.7571	3.0649	0.1643	3.014	-0.7428	-0.9781	0.8526	0.0012	0.0601	-0.7726	
9	-0.5794	-0.0445	3.44	-0.7666	0.8668	2.3112	1.1003	-1.4491	-3.719	0.0492	-6.8411	
10	0.5312	-0.0919	-1.3187	0.4284	2.6753	2.1158	-0.7679	-0.3967	-0.5584	0.0727	1.664	

w_1: Weights vector between input and the hidden layer of the neural network.

& w_2: Weights vector between the hidden and the output layer of the neural network.

b_1: Bias vector between input and the hidden layer of the neural network.

& b_2: Bias between the hidden and the output layer of the neural network.

The normalized empirical equation for the ANN model with nine inputs as follows:

$$(\text{Water Cut})_{calc} = 0.84259^* (\text{Water Cut})_{actual} + 6.0223 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6.5)$$

Figures 6.39 and 6.40 represent the overall results for low GOR and high GOR wells, respectively. Average Absolute Relative Percentage Error for low GOR wells is 9.2% while error for high GOR wells is 7.99%. Both give an error less than the acceptable limit of 10% which confirm ANN model accuracy with nine inputs.

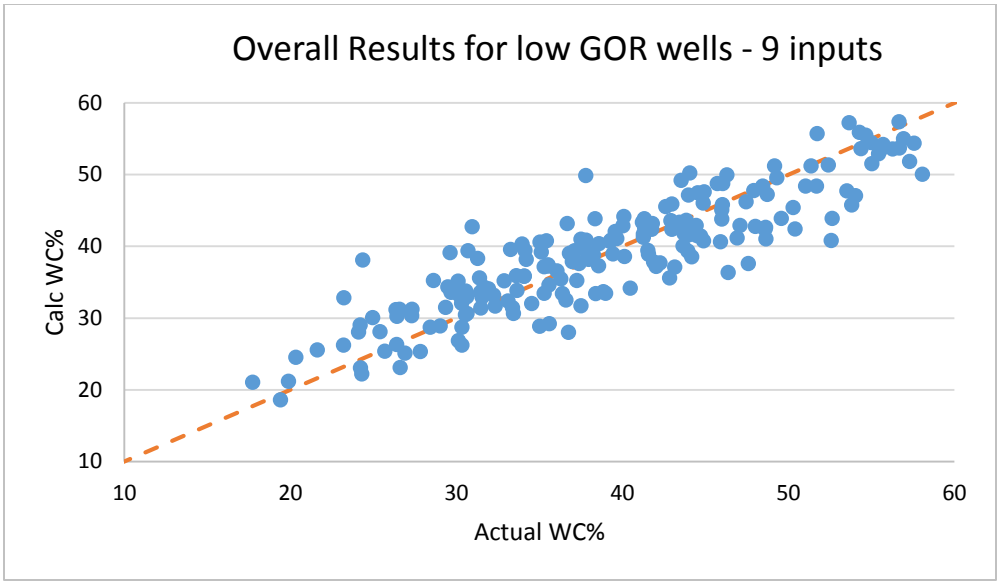


Figure 6.39 Overall results with the ANN model of nine inputs for low GOR wells

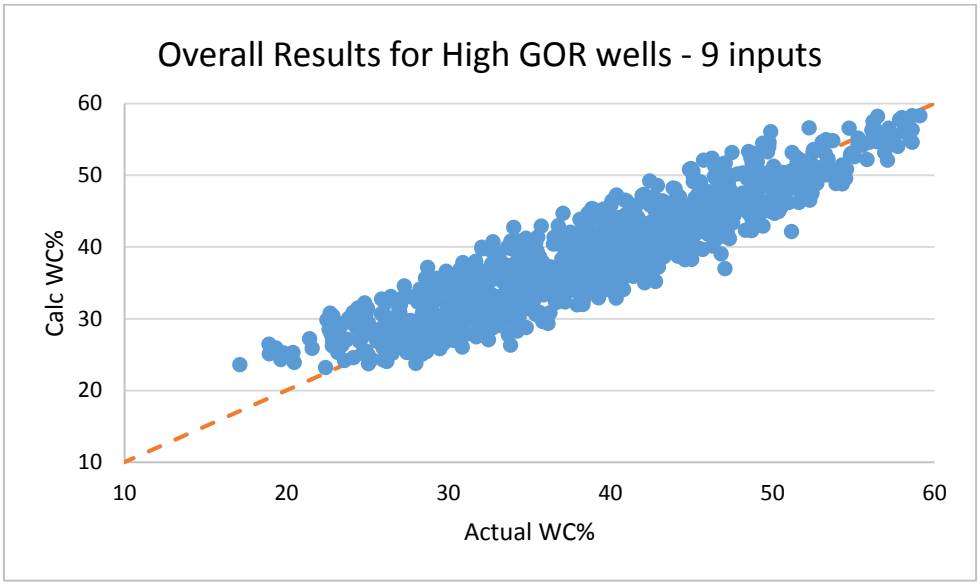


Figure 6.40 Overall results with the ANN model of nine inputs for high GOR wells

To sum up, developed empirical correlation works for low GOR applications, and ANN model applies to all range of GOR.

IV. Summarized Results

Table 6.13 summarizes all results from available multiphase flow correlations and developed models. All selected multiphase flow correlations and the new developed empirical correlation failed to estimate the water-cut accurately within the acceptable error for high GOR wells. The developed empirical correlation achieved an error of less than 10% in low GOR wells. The ANN model was successful in achieving the research objective with an error less than 10% covering the whole range of GOR. Failure of the empirical correlation to estimate water-cut in high GOR wells was due to inapplicability of non-linear multiple regression to dictate the relationship between inputs and output, unlike ANN as neurons will search for the best fit between inputs and output by weights alteration. This process is achieved via training and testing of data.

Table 6.13 Summarized Table of Errors per Correlations and Models

Method/Model		E_{ar}% in WC		
		GOR<2,000	ALL	GOR>=2,000
Multiphase flow correlation	Guo Ghalambor	13	27	39
	Hagedorn Brown	12	23	33
	Beggs & Brill	16	32	46
	Gray	11	21	29
	Duns & Ros	14	26	36
	Average	13	25	37
Non-linear multiple regression		8.3	20.3	25.1
Artificial Neural Network with five inputs		7.2	8.13	8.4
Artificial Neural Network with nine inputs		9.2	8.18	7.99

6.4 ANN Model Verification

The ANN optimized model was also verified with another set of data of 430 data points using the same weights and biases in section 6.3.2 (III.i). The obtained error (E_{ar}%) was 7.72% overall, which reflected model robustness. **Figure 6.41** shows the relationship between actual water-cut values and the calculated values with a high coefficient of determination of 0.92.

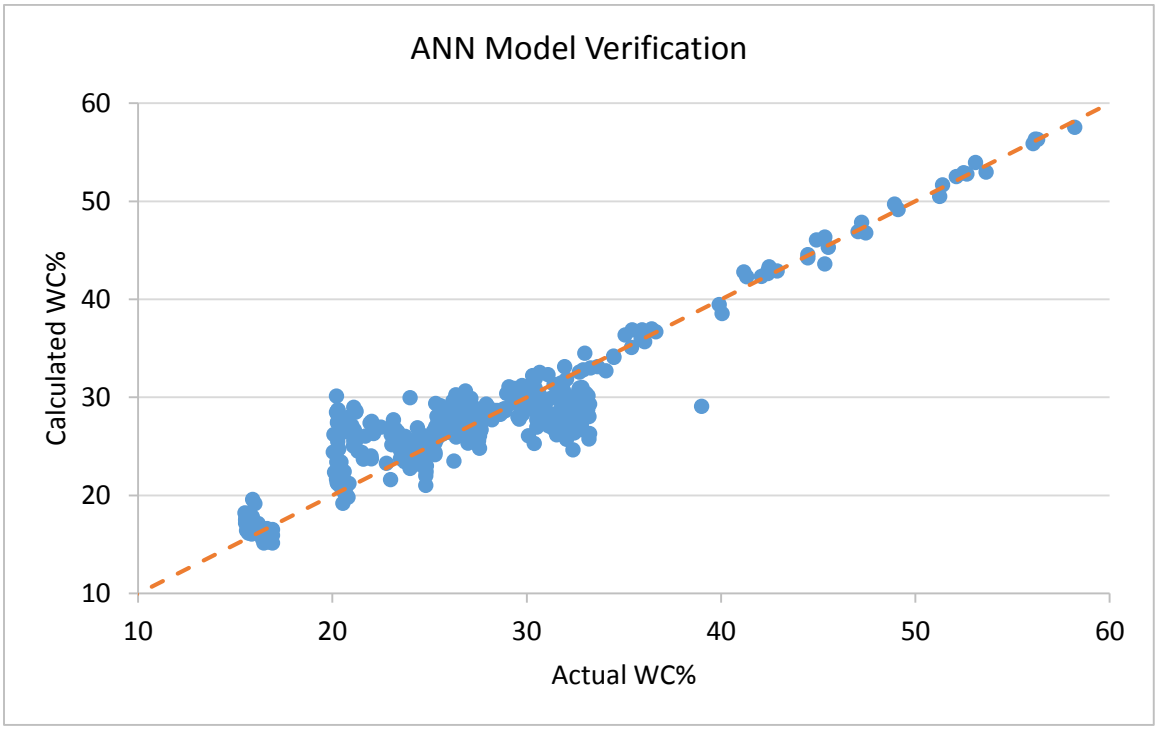


Figure 6.41 Overall results for ANN model verification

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

The following points resulted from this research:

- A new model was developed to estimate high water cut in high GOR wells.
- Evaluation of existing multi-phase flow correlations revealed their inapplicability to estimate water-cut in high GOR wells.
- The developed empirical correlation was found applicable in low GOR wells only.
- The developed ANN model was found applicable to all range of GOR with the least errors.
- The five selected input parameters were sufficient to produce accurate models, and the eliminated four inputs had no significant improvement on the results.
- The developed ANN model was also validated against a new set of data.
- The ANN technique proves its high accuracy for estimating water cut with a minimum error with a provided dataset where it can be used as an alternative tool to estimate water cut when measurement devices such as separators go down or require maintenance.

7.2 Recommendations for future work

An accurate model was developed using ANN technique. There are potential areas of research that could be explored to extend this research further:

- Validate the developed ANN model against different datasets from other fields with high GOR.
- Evaluate the model against different types of wells (vertical/deviated/horizontal and multilateral wells).
- Assess the impact of cost reduction on field operations by reducing the dependency on metering technologies.

APPENDIX A

LIQUID RATE CALCULATIONS USING PROSPER™

PROSPER system calculations snapshots

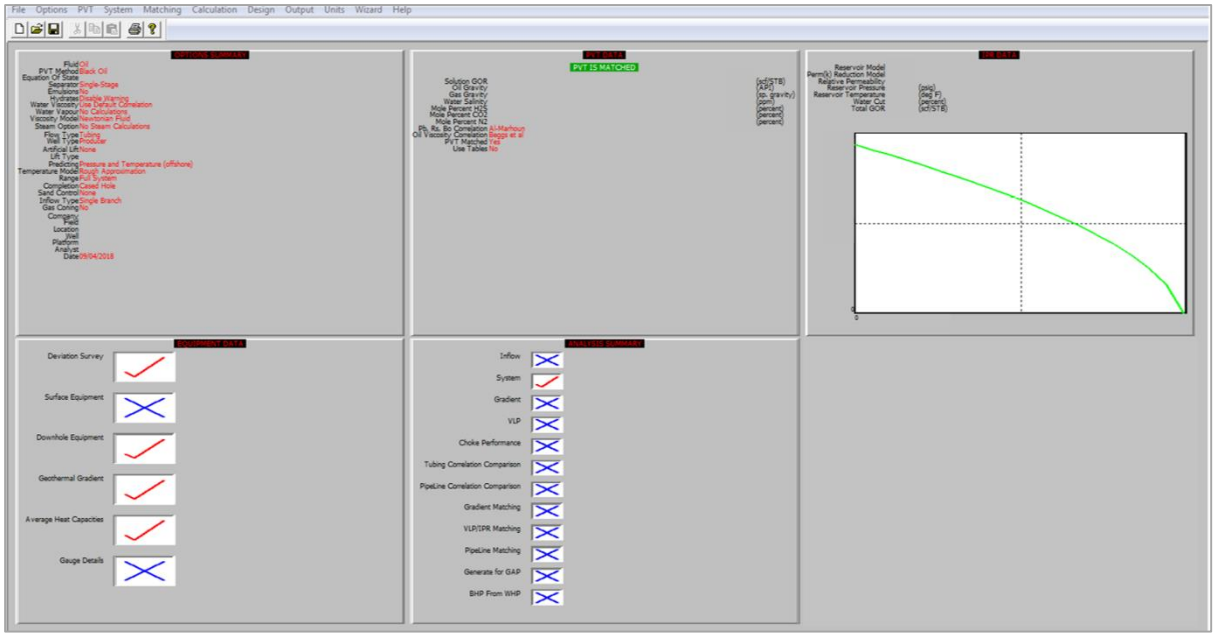
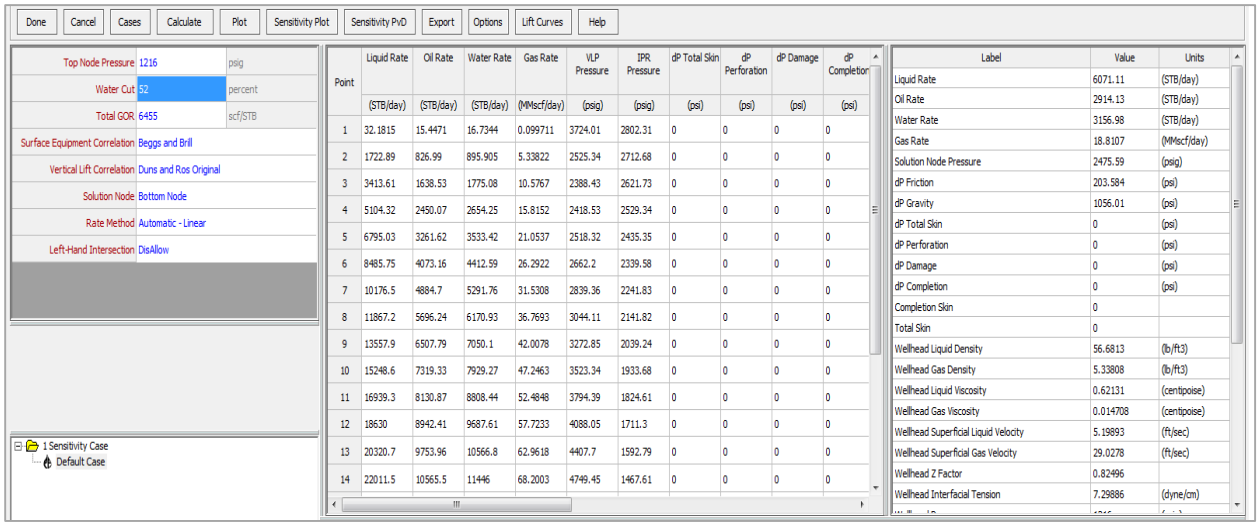


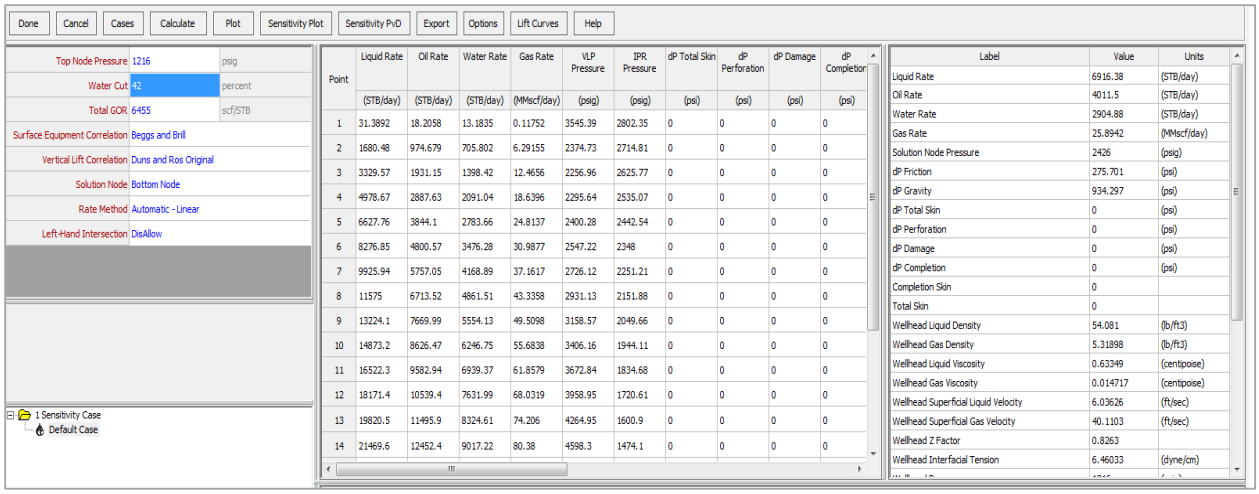
Figure A.1 PROSPER startup interface



Point	Liquid Rate	Oil Rate	Water Rate	Gas Rate	VLP Pressure	IPR Pressure	dP Total Skin	dP Perforation	dP Damage	dP Completion
	(STB/day)	(STB/day)	(STB/day)	(MMscf/day)	(psig)	(psig)	(psi)	(psi)	(psi)	(psi)
1	32.1815	15.4471	16.7344	0.099711	3724.01	2802.31	0	0	0	0
2	1722.89	826.99	895.905	5.33822	2525.34	2712.68	0	0	0	0
3	3413.61	1638.53	1775.08	10.5767	2388.43	2621.73	0	0	0	0
4	5104.32	2450.07	2654.25	15.8152	2418.53	2529.34	0	0	0	0
5	6795.03	3261.62	3533.42	21.0537	2518.32	2435.35	0	0	0	0
6	8485.75	4073.16	4412.59	26.2922	2662.2	2339.58	0	0	0	0
7	10176.5	4884.7	5291.76	31.5308	2839.36	2241.83	0	0	0	0
8	11867.2	5696.24	6170.93	36.7693	3044.11	2141.82	0	0	0	0
9	13557.9	6507.79	7050.1	42.0078	3272.85	2039.24	0	0	0	0
10	15248.6	7319.33	7929.27	47.2463	3523.34	1933.68	0	0	0	0
11	16939.3	8130.87	8808.44	52.4848	3794.39	1824.61	0	0	0	0
12	18630	8942.41	9687.61	57.7233	4088.05	1711.3	0	0	0	0
13	20320.7	9753.96	10566.8	62.9618	4407.7	1592.79	0	0	0	0
14	22011.5	10565.5	11446	68.2003	4749.45	1467.61	0	0	0	0

Label	Value	Units
Liquid Rate	6071.11	(STB/day)
Oil Rate	2914.13	(STB/day)
Water Rate	3156.98	(STB/day)
Gas Rate	18.8107	(MMscf/day)
Solution Node Pressure	2475.59	(psig)
dP Friction	203.584	(psi)
dP Gravity	1056.01	(psi)
dP Total Skin	0	(psi)
dP Perforation	0	(psi)
dP Damage	0	(psi)
dP Completion	0	(psi)
Completion Skin	0	
Total Skin	0	
Wellhead Liquid Density	56.6813	(lb/ft3)
Wellhead Gas Density	5.33808	(lb/ft3)
Wellhead Liquid Viscosity	0.62131	(centipoise)
Wellhead Gas Viscosity	0.014708	(centipoise)
Wellhead Superficial Liquid Velocity	5.19893	(ft/sec)
Wellhead Superficial Gas Velocity	29.0278	(ft/sec)
Wellhead Z Factor	0.82496	
Wellhead Interfacial Tension	7.29886	(dyne/cm)

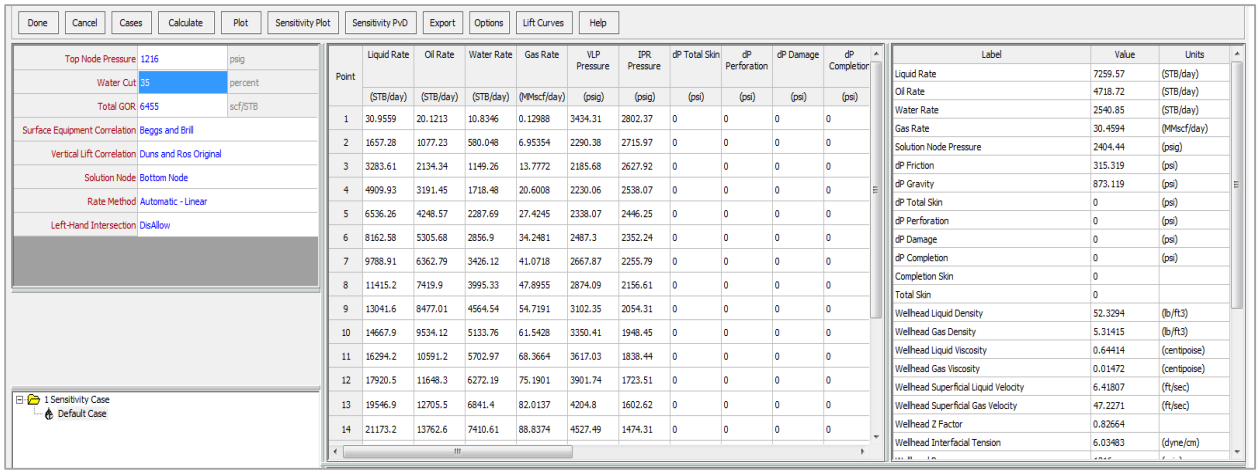
Figure A.2: PROSPER system calculation; initial trial



Point	Liquid Rate (STB/day)	Oil Rate (STB/day)	Water Rate (STB/day)	Gas Rate (MMscf/day)	VLP Pressure (psig)	IPR Pressure (psig)	dP Total Skin (psi)	dP Perforation (psi)	dP Damage (psi)	dP Completion (psi)
1	31.3892	18.2058	13.1835	0.11752	3545.39	2802.35	0	0	0	0
2	1680.48	974.679	705.802	6.29155	2374.73	2714.81	0	0	0	0
3	3329.57	1931.15	1398.42	12.4656	2256.96	2625.77	0	0	0	0
4	4978.67	2887.63	2091.04	18.6396	2295.64	2535.07	0	0	0	0
5	6627.76	3844.1	2783.66	24.8137	2400.28	2442.54	0	0	0	0
6	8276.85	4800.57	3476.28	30.9877	2547.22	2348	0	0	0	0
7	9925.94	5757.05	4168.89	37.1617	2726.12	2251.21	0	0	0	0
8	11575	6713.52	4861.51	43.3358	2931.13	2151.88	0	0	0	0
9	13224.1	7669.99	5554.13	49.5098	3158.57	2049.66	0	0	0	0
10	14873.2	8626.47	6246.75	55.6838	3406.16	1944.11	0	0	0	0
11	16522.3	9582.94	6939.37	61.8579	3672.84	1834.68	0	0	0	0
12	18171.4	10539.4	7631.99	68.0319	3958.95	1720.61	0	0	0	0
13	19820.5	11495.9	8324.61	74.206	4264.95	1600.9	0	0	0	0
14	21469.6	12452.4	9017.22	80.38	4598.3	1474.1	0	0	0	0

Label	Value	Units
Liquid Rate	6916.38	(STB/day)
Oil Rate	4011.5	(STB/day)
Water Rate	2904.88	(STB/day)
Gas Rate	25.8942	(MMscf/day)
Solution Node Pressure	2426	(psig)
dP Friction	275.701	(psi)
dP Gravity	934.297	(psi)
dP Total Skin	0	(psi)
dP Perforation	0	(psi)
dP Damage	0	(psi)
dP Completion	0	(psi)
Completion Skin	0	
Total Skin	0	
Wellhead Liquid Density	54.081	(lb/ft3)
Wellhead Gas Density	5.31898	(lb/ft3)
Wellhead Liquid Viscosity	0.63349	(centipoise)
Wellhead Gas Viscosity	0.014717	(centipoise)
Wellhead Superficial Liquid Velocity	6.03626	(ft/sec)
Wellhead Superficial Gas Velocity	40.1103	(ft/sec)
Wellhead Z Factor	0.8263	
Wellhead Interfacial Tension	6.46033	(dyne/cm)

Figure A.3: PROSPER system calculation; the first trial



Point	Liquid Rate (STB/day)	Oil Rate (STB/day)	Water Rate (STB/day)	Gas Rate (MMscf/day)	VLP Pressure (psig)	IPR Pressure (psig)	dP Total Skin (psi)	dP Perforation (psi)	dP Damage (psi)	dP Completion (psi)
1	30.9559	20.1213	10.8346	0.12988	3434.31	2802.37	0	0	0	0
2	1657.28	1077.23	580.048	6.95354	2290.38	2715.97	0	0	0	0
3	3283.61	2134.34	1149.26	13.7772	2185.68	2627.92	0	0	0	0
4	4909.93	3191.45	1718.48	20.6008	2230.06	2538.07	0	0	0	0
5	6536.26	4248.57	2287.69	27.4245	2338.07	2446.25	0	0	0	0
6	8162.58	5305.68	2856.9	34.2481	2487.3	2352.24	0	0	0	0
7	9788.91	6362.79	3426.12	41.0718	2667.87	2255.79	0	0	0	0
8	11415.2	7419.9	3995.33	47.8955	2874.09	2156.61	0	0	0	0
9	13041.6	8477.01	4564.54	54.7191	3102.35	2054.31	0	0	0	0
10	14667.9	9534.12	5133.76	61.5428	3350.41	1948.45	0	0	0	0
11	16294.2	10591.2	5702.97	68.3664	3617.03	1838.44	0	0	0	0
12	17920.5	11648.3	6272.19	75.1901	3901.74	1723.51	0	0	0	0
13	19546.9	12705.5	6841.4	82.0137	4204.8	1602.62	0	0	0	0
14	21173.2	13762.6	7410.61	88.8374	4527.49	1474.31	0	0	0	0

Label	Value	Units
Liquid Rate	7259.57	(STB/day)
Oil Rate	4718.72	(STB/day)
Water Rate	2540.85	(STB/day)
Gas Rate	30.4594	(MMscf/day)
Solution Node Pressure	2404.44	(psig)
dP Friction	315.319	(psi)
dP Gravity	873.119	(psi)
dP Total Skin	0	(psi)
dP Perforation	0	(psi)
dP Damage	0	(psi)
dP Completion	0	(psi)
Completion Skin	0	
Total Skin	0	
Wellhead Liquid Density	52.3294	(lb/ft3)
Wellhead Gas Density	5.31415	(lb/ft3)
Wellhead Liquid Viscosity	0.64414	(centipoise)
Wellhead Gas Viscosity	0.01472	(centipoise)
Wellhead Superficial Liquid Velocity	6.41807	(ft/sec)
Wellhead Superficial Gas Velocity	47.2271	(ft/sec)
Wellhead Z Factor	0.82664	
Wellhead Interfacial Tension	6.03483	(dyne/cm)

Figure A.4: PROSPER system calculation; second trial

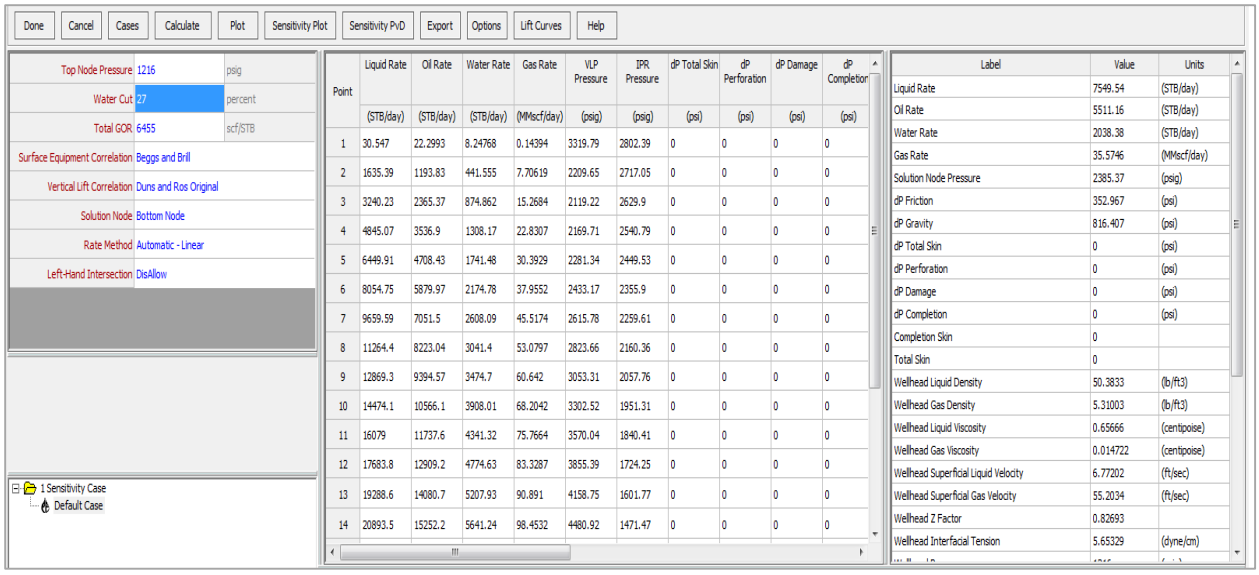


Figure A.5: PROSPER system calculation; last trial

Example of matching the liquid rate of test separator and from the multiphase correlation (Actual WC=52, Act Q_L= 7952)

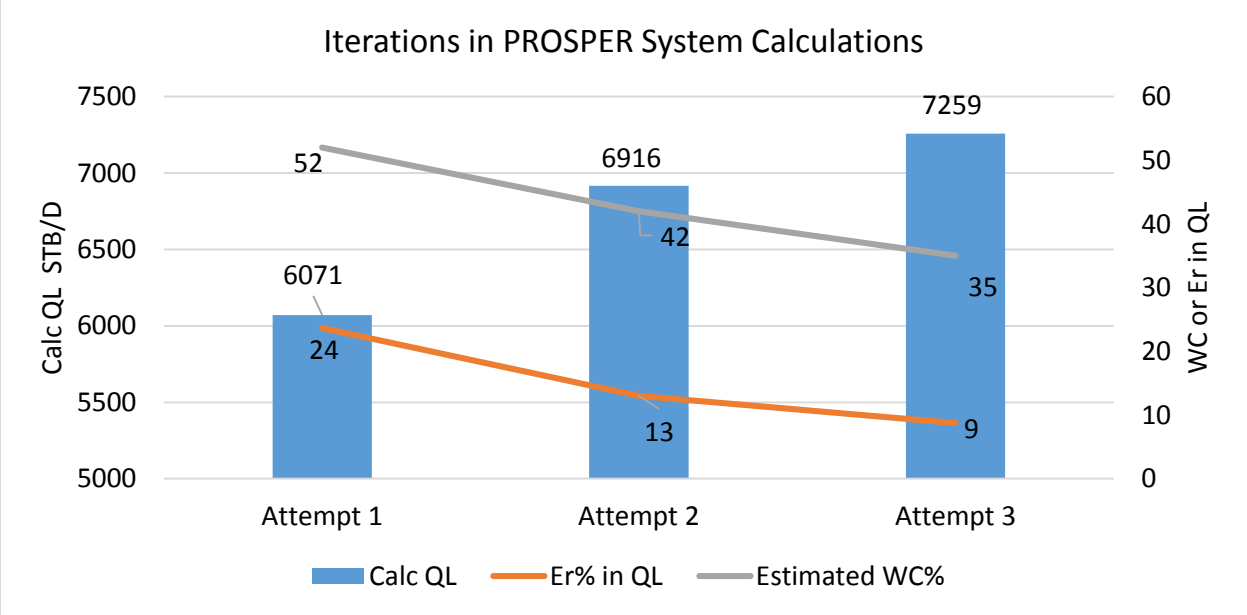


Figure A.6: PROSPER liquid rate iterations summary

APPENDIX B

ARTIFICIAL NEURAL NETWORKS (ANN) OPTIMIZATION SUMMARY

I. Optimizing training functions (Table 4.1)

Table B.1 Most common Training (or Testing) Functions used in ANN

Type#	Name Description	MATLAB function	# of neurons in hidden layer 1	# of neurons in hidden layer 2
0	Levenberg-Marquardt with one hidden layer	trainlm	10	
1	Levenberg-Marquardt with two hidden layers	trainlm	10	10
2	Backpropagation using gradient descent	traingd	10	10
3	Backpropagation using gradient descent with momentum and adaptive (variable) learning rate backpropagation	traingdx	10	10
4	Backpropagation using Broyden, Fletcher, Goldfarb, and Shanno (BFGS) Quasi-Newton	trainbfg	10	10
5	Backpropagation using scaled conjugate gradient	trainscg	10	10
6	Backpropagation with One Step Secant	trainoss	10	10
7	Resilient backpropagation	trainrp	10	10
8	Batch backpropagation	trainb	10	10

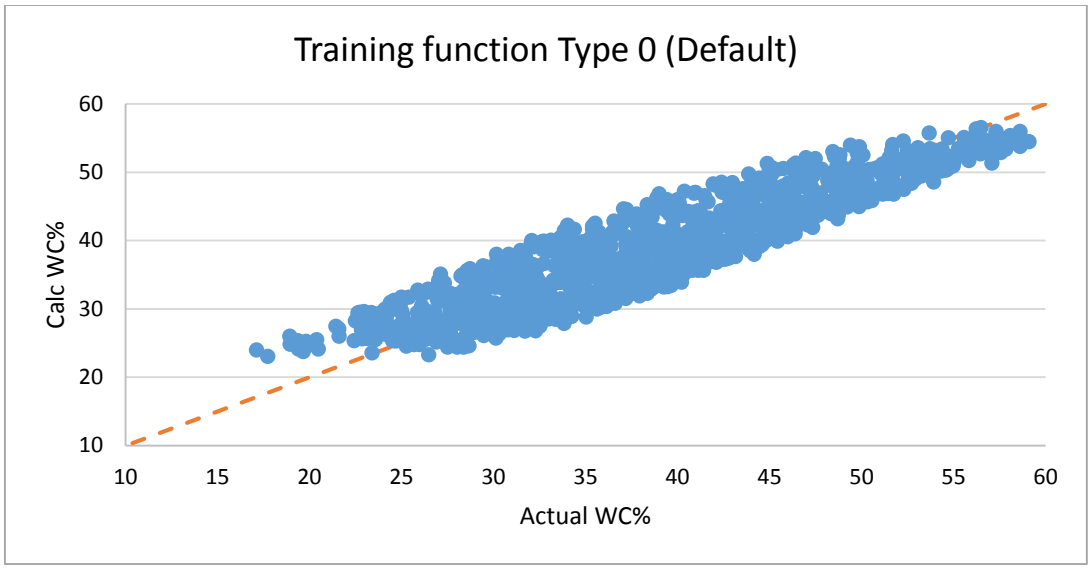


Figure B.1 Type = 0 using Levenberg-Marquardt (trainlm) algorithm with 1 hidden layer

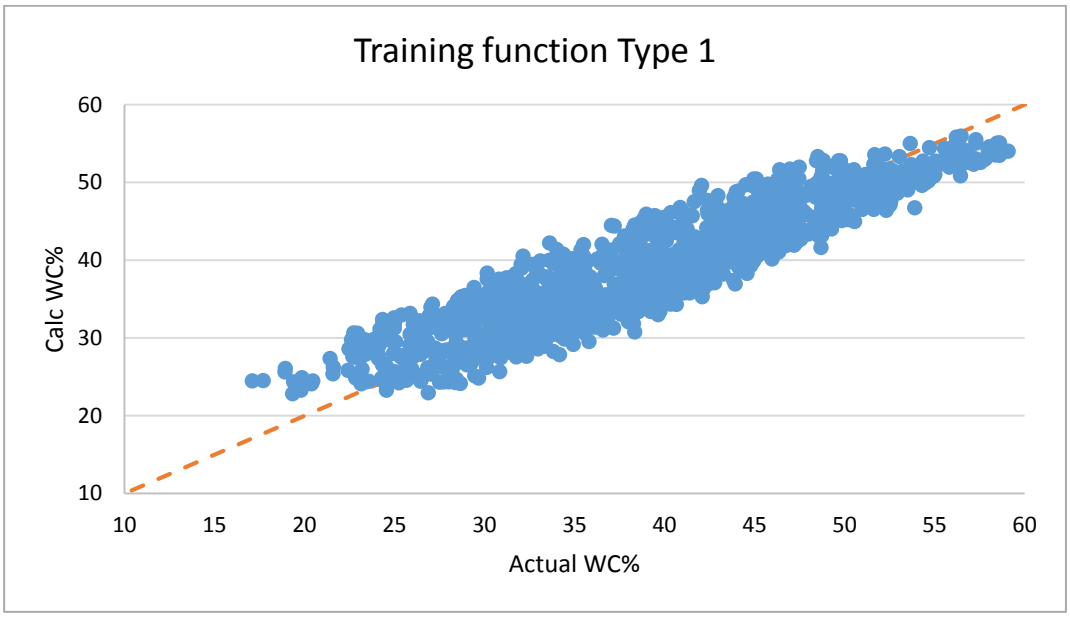


Figure B.2 Type = 1 using Levenberg-Marquardt (trainlm) algorithm with 2 hidden layers

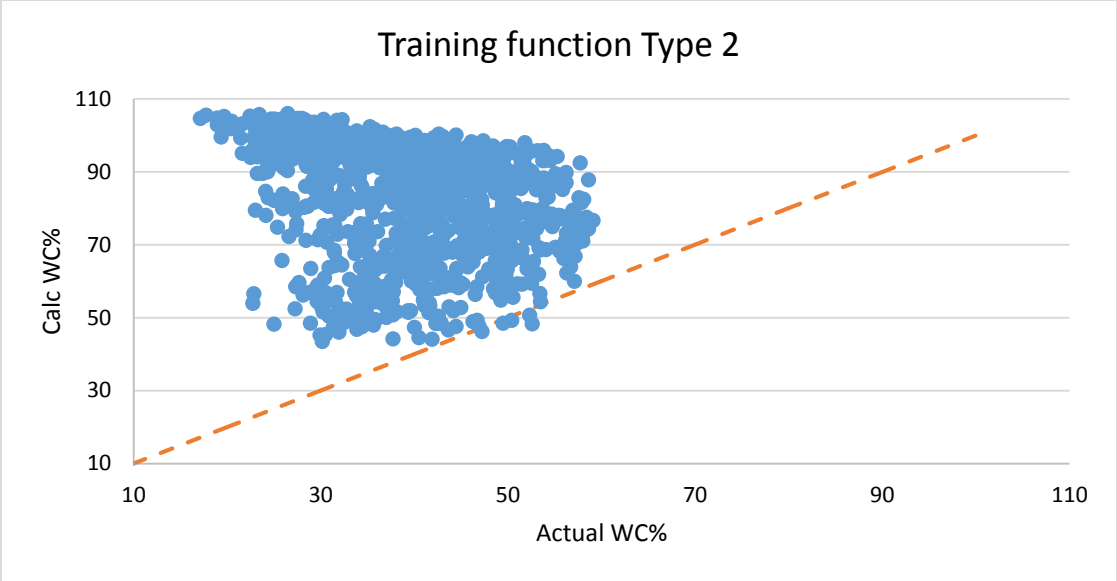


Figure B.3 Type = 2 using gradient descent (traingd) algorithm

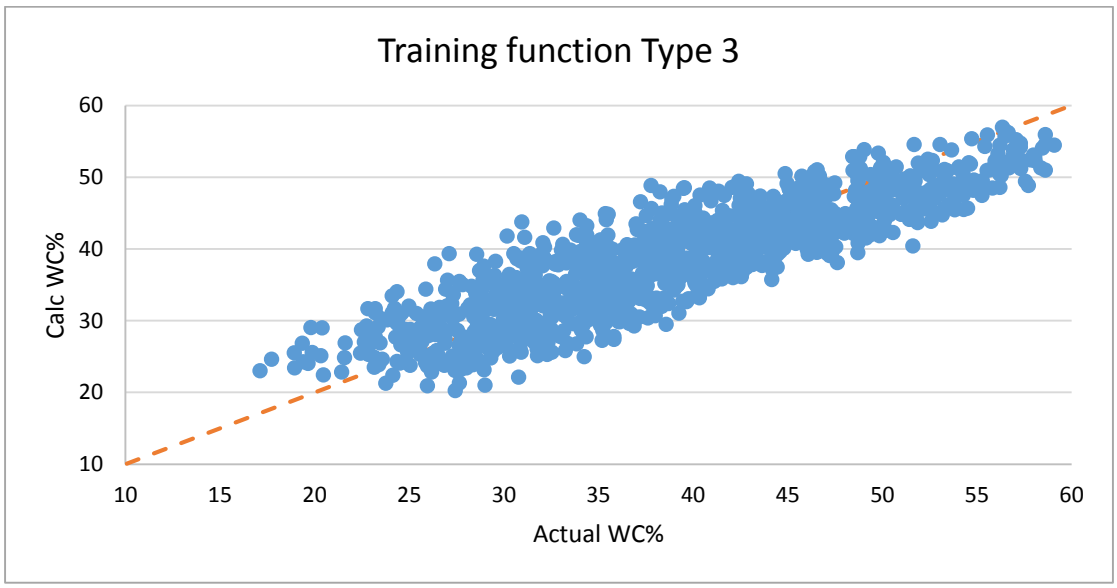


Figure B.4 Type = 3 using gradient descent with momentum (traindx) algorithm

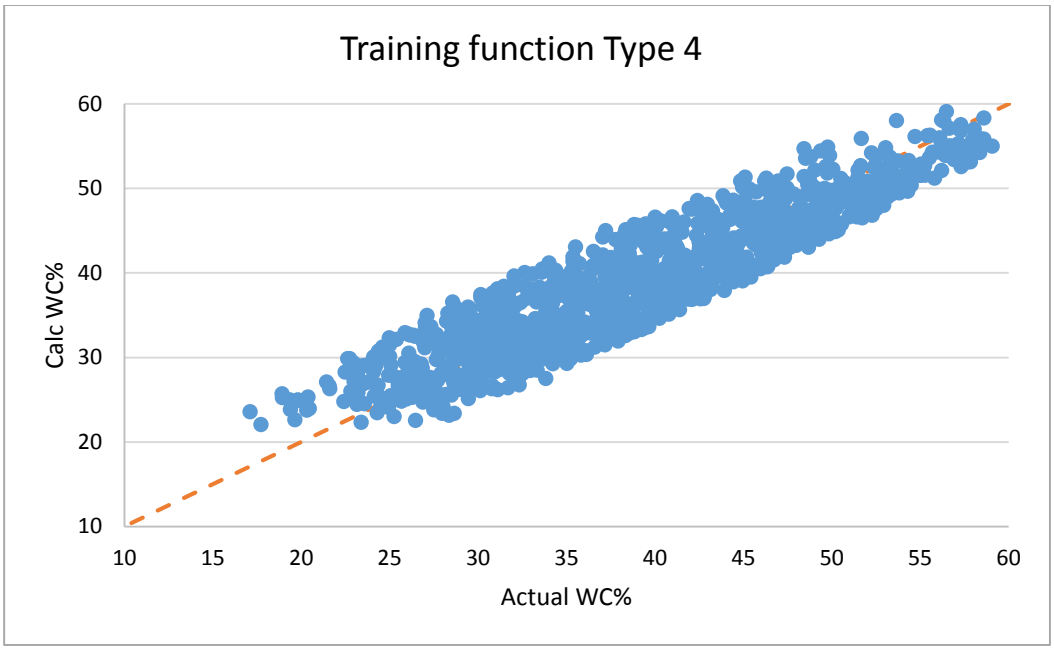


Figure B.5 Type = 4 using BFGS Quasi-Newton (trainbfg) algorithm

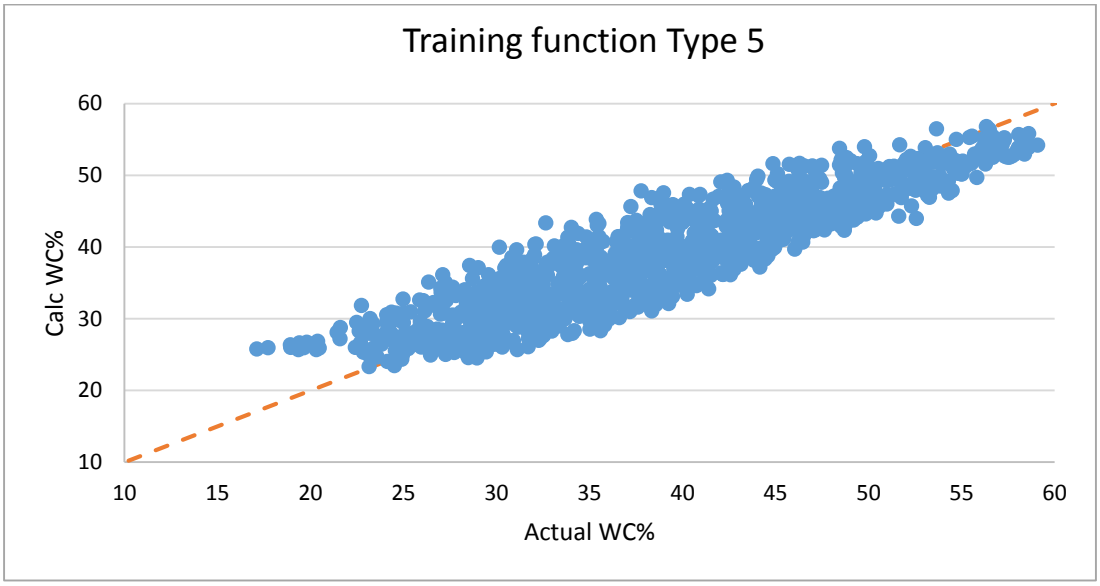


Figure B.6 Type = 5 using scaled conjugate gradient (trainscg) algorithm

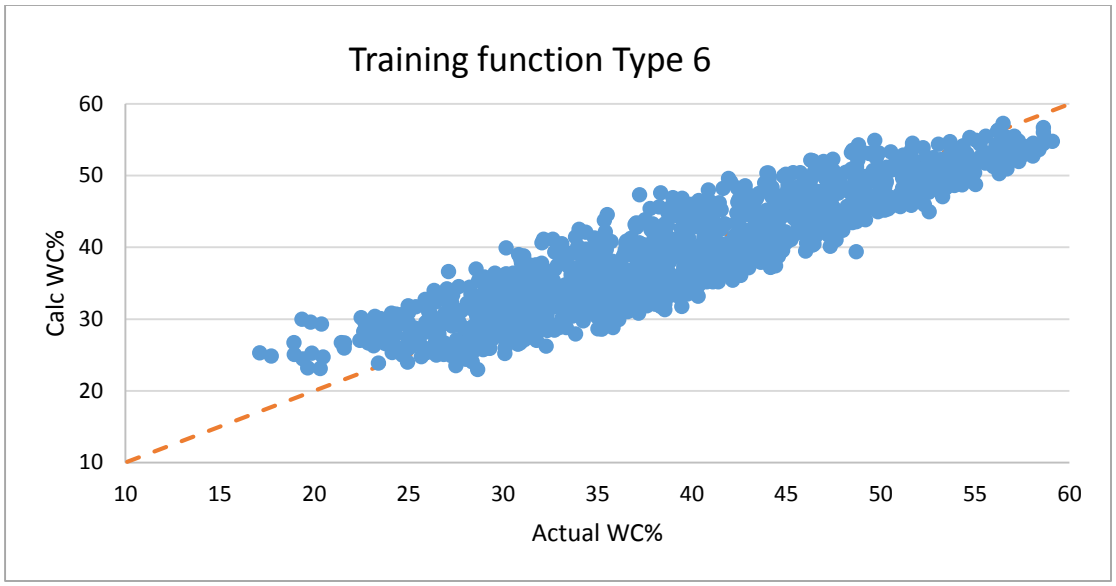


Figure B.7 Type = 6 using One Step Secant (trainoss) algorithm

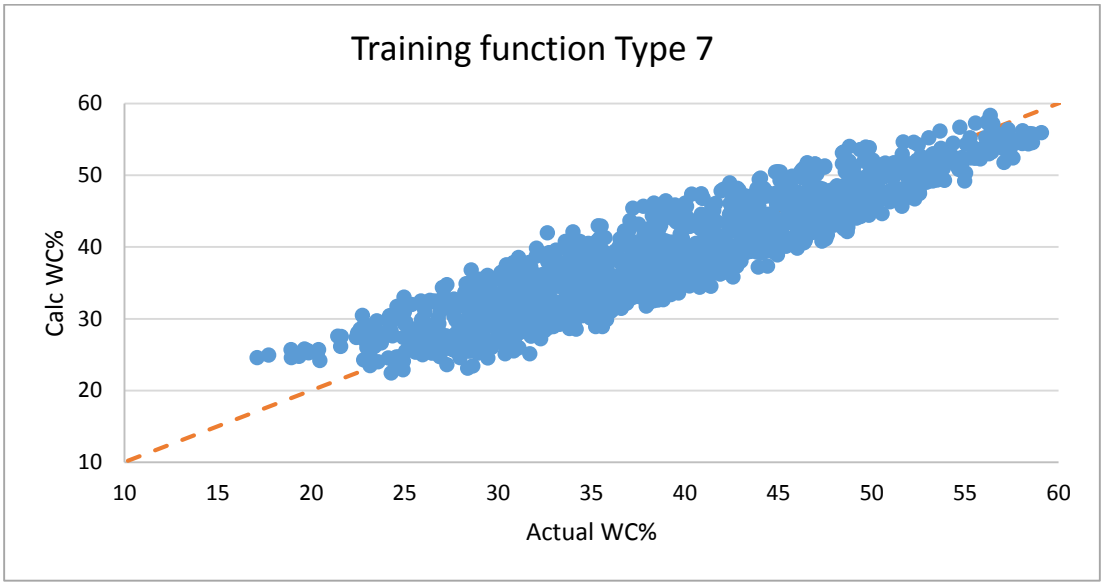


Figure B.8 Type = 7 using resilient backpropagation (trainrp) algorithm

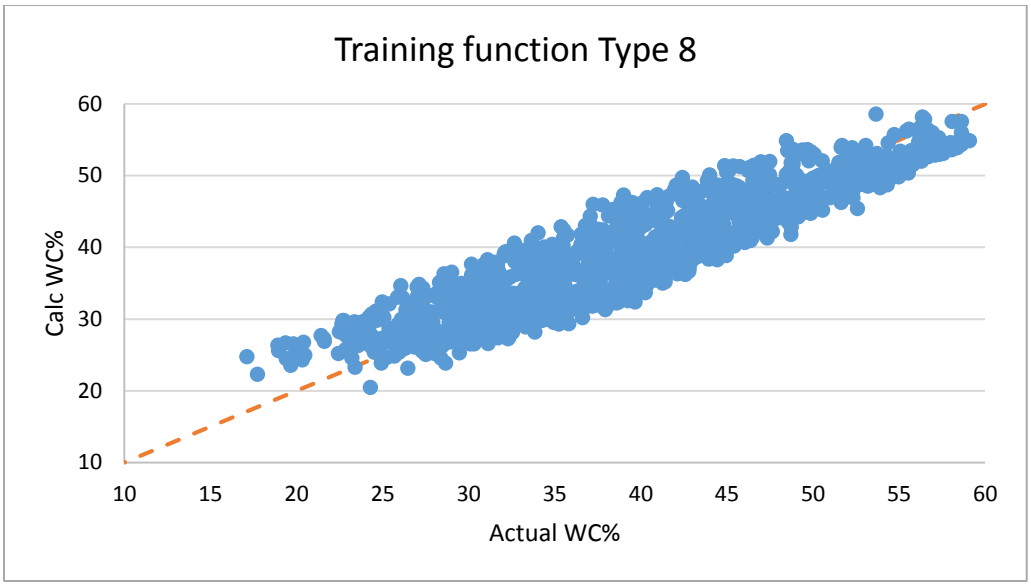


Figure B.9 Type = 8 batch backpropagation (trainb) algorithm

II. Optimizing transfer function (Table 4.2)

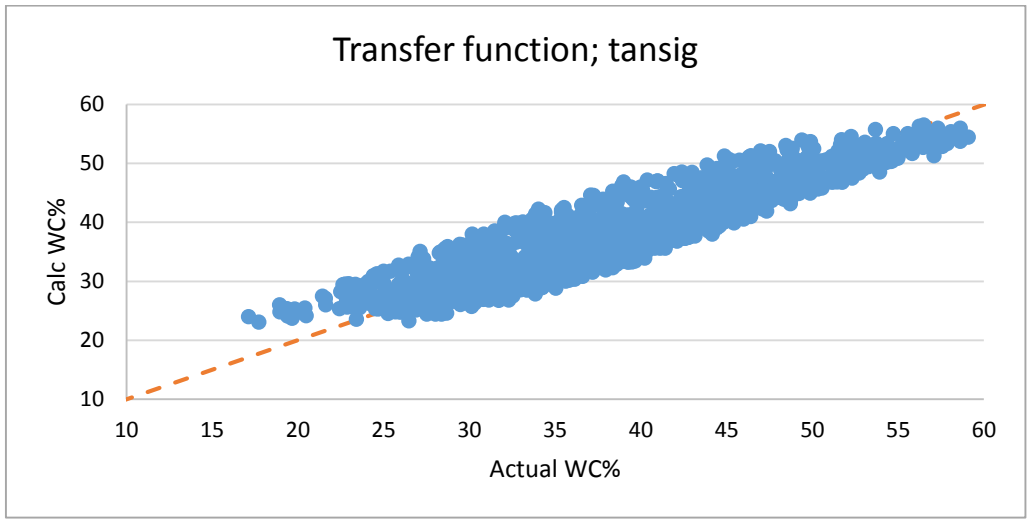


Figure B.10 Hyperbolic tangent sigmoid transfer function

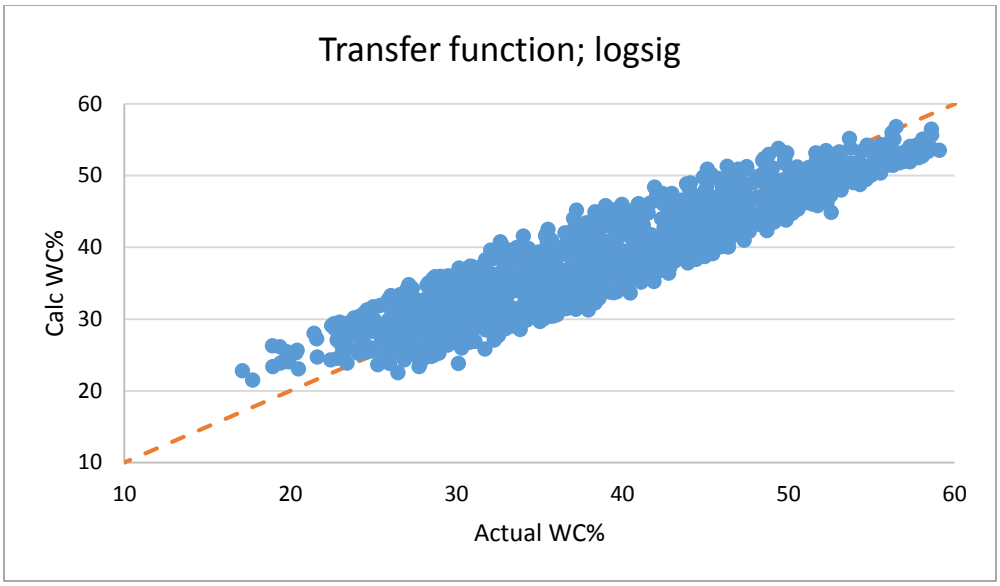


Figure B.11 Log sigmoid transfer function

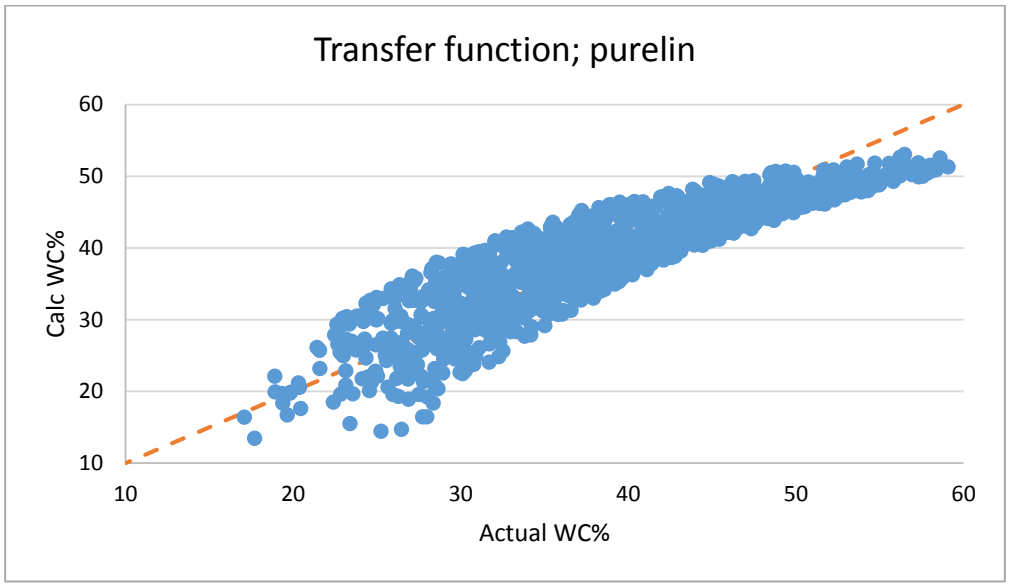


Figure B.12 Pure linear transfer function

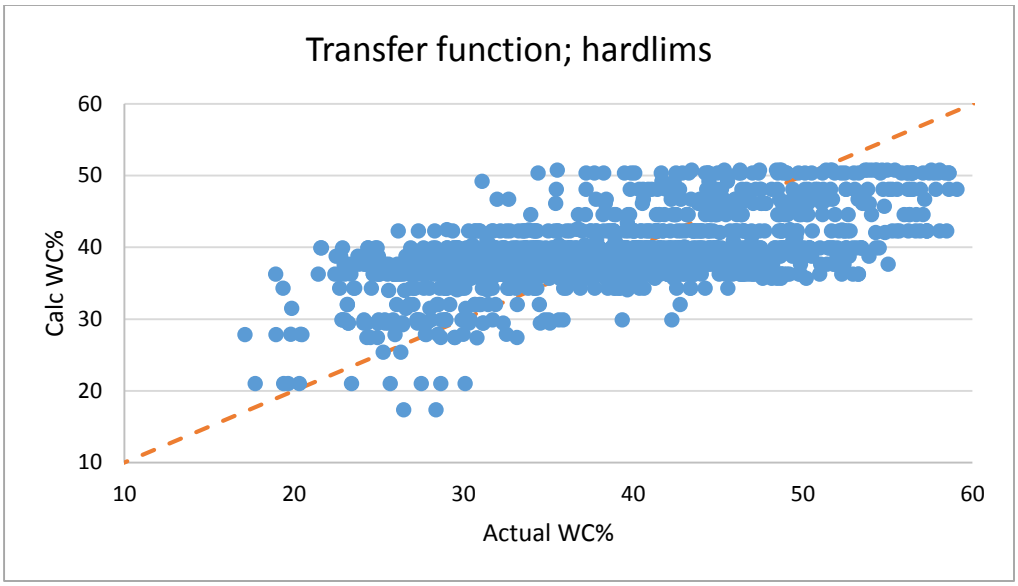


Figure B.13 Symmetric hard limit transfer function (not differentiable)

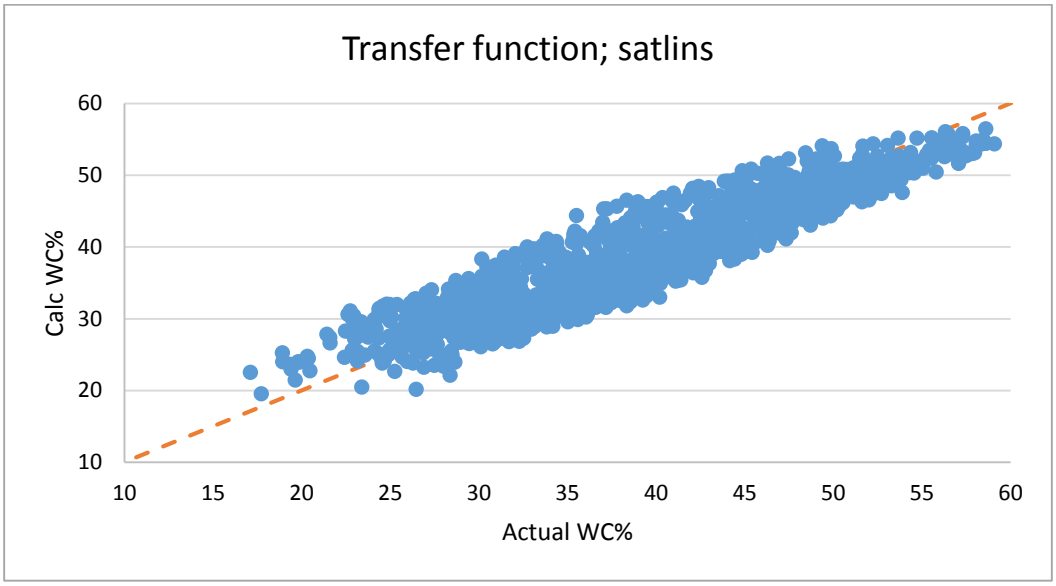


Figure B.14 Symmetric saturated linear transfer function

V. Optimizing the number of neurons

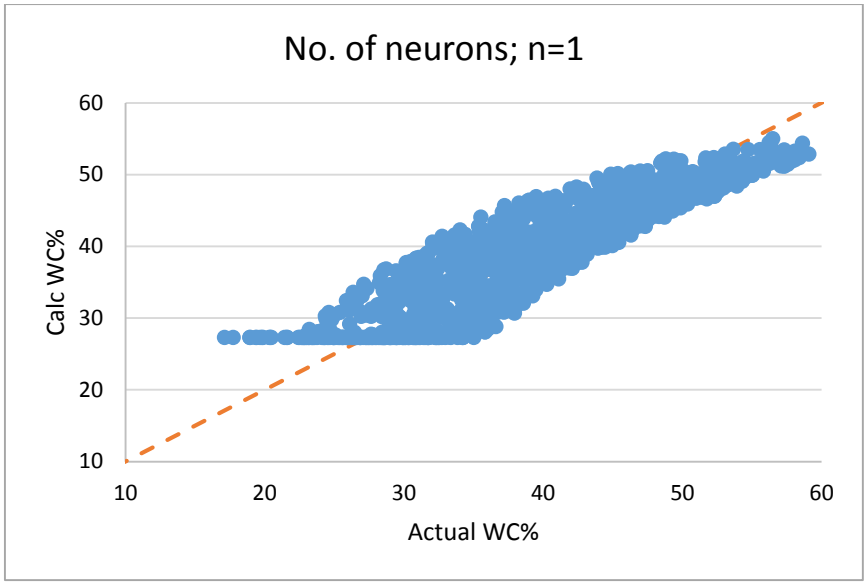


Figure B.15 Symmetric saturating linear (satlins) transfer function (n=1)

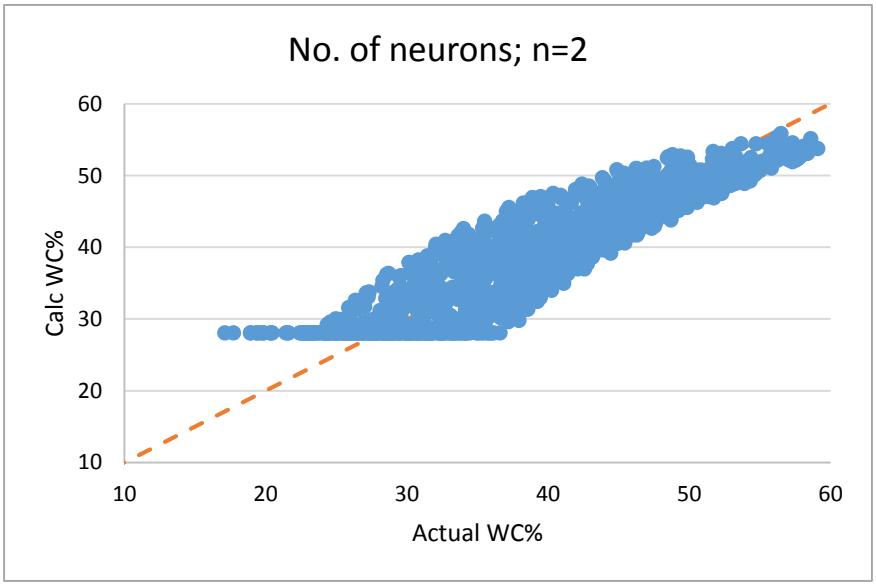


Figure B.16 Symmetric saturating linear (satlins) transfer function (n=2)

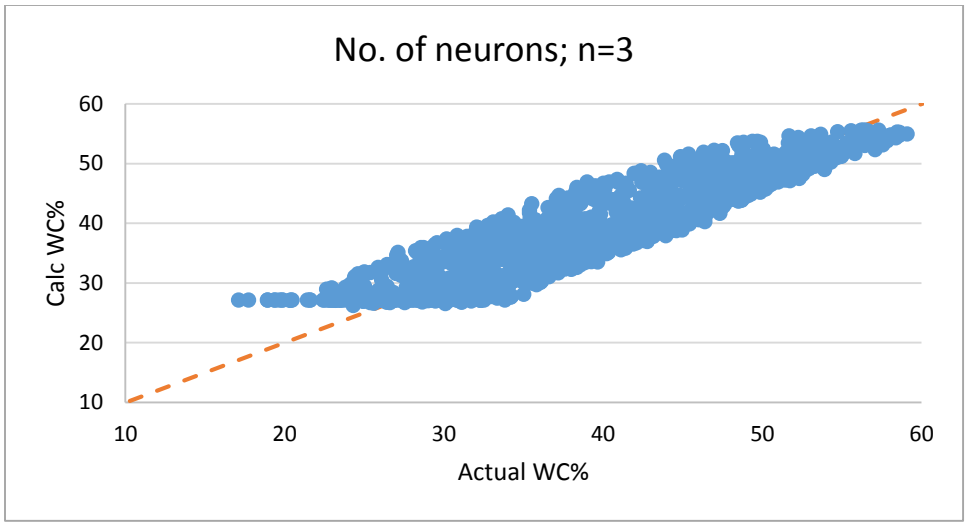


Figure B.17 Symmetric saturating linear (satlins) transfer function (n=3)

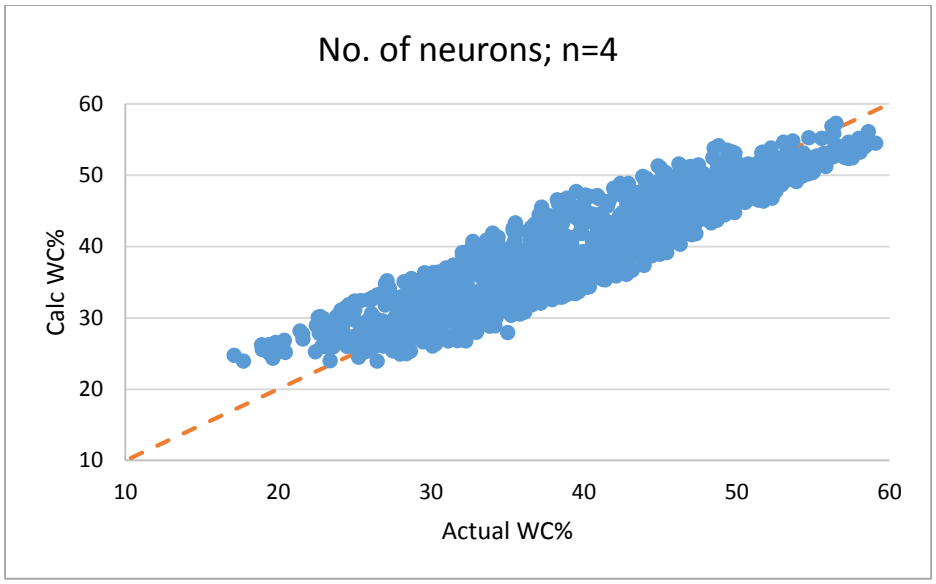


Figure B.18 Symmetric saturating linear (satlins) transfer function (n=4)

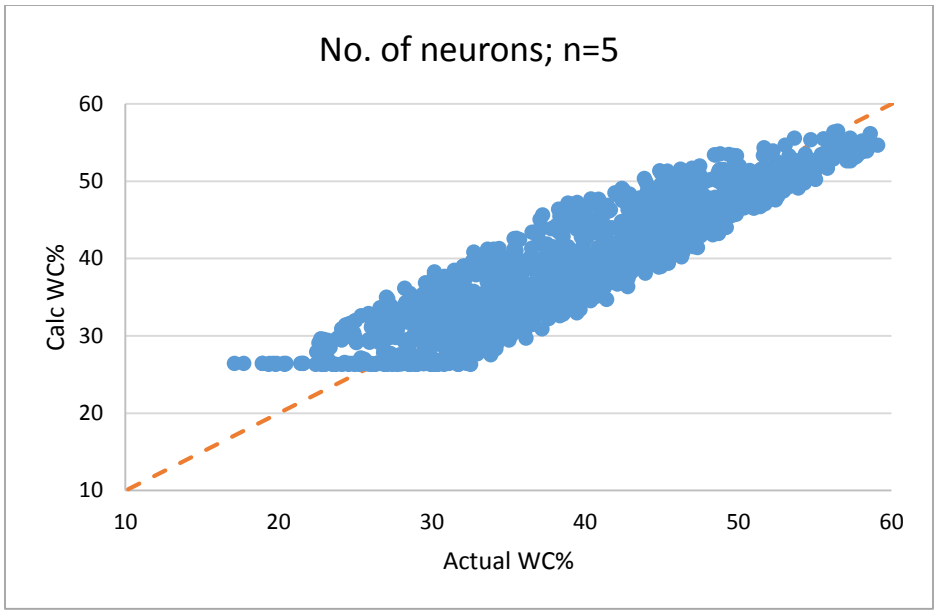


Figure B.19 Symmetric saturating linear (satlins) transfer function (n=5)

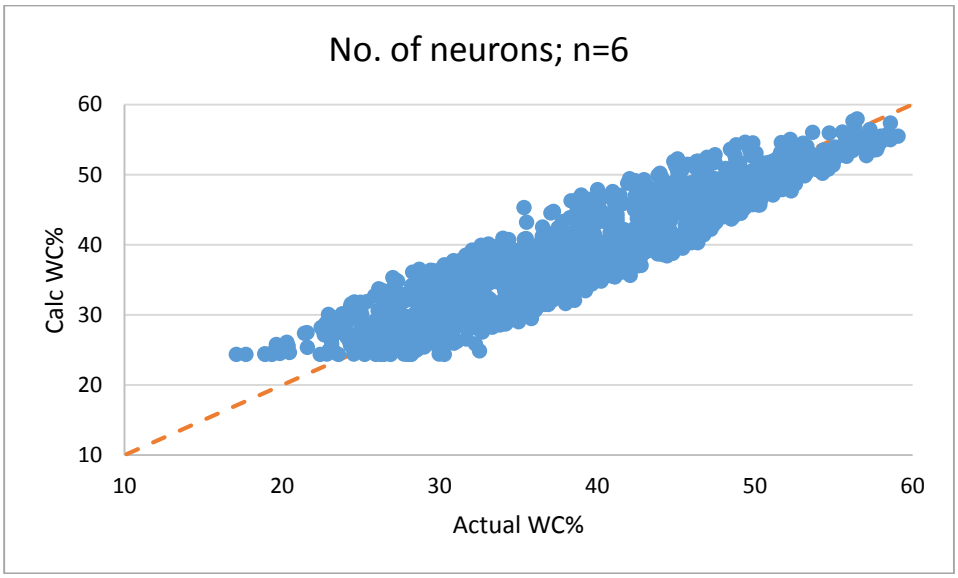


Figure B.20 Symmetric saturating linear (satlins) transfer function (n=6)

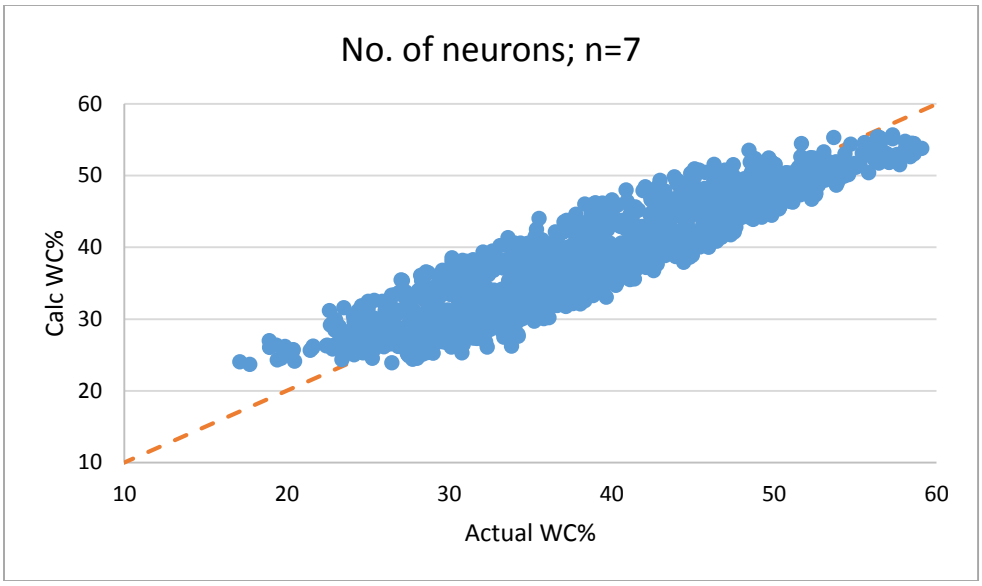


Figure B.21 Symmetric saturating linear (satlins) transfer function (n=7)

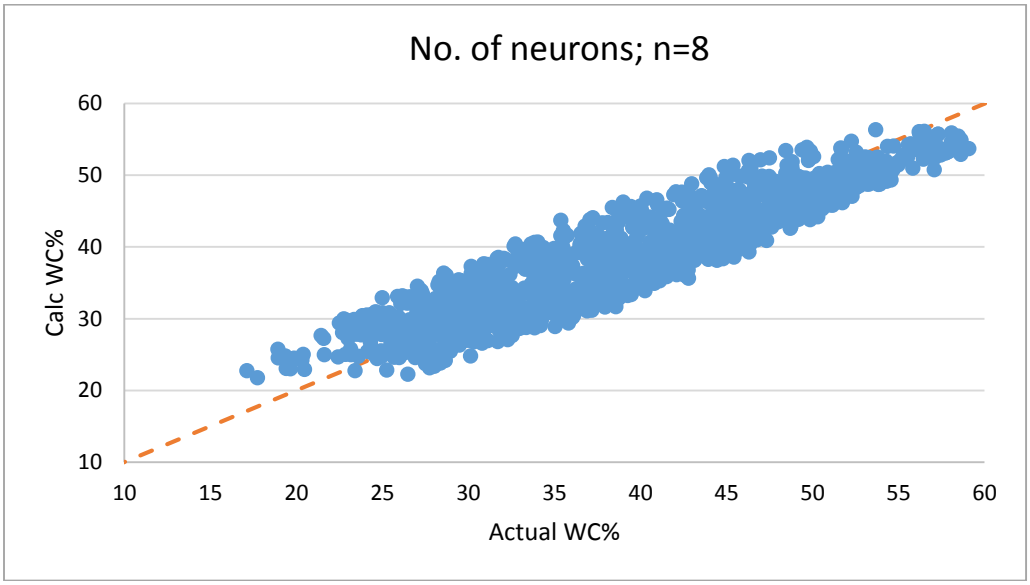


Figure B.22 Symmetric saturating linear (satlins) transfer function (n=8)

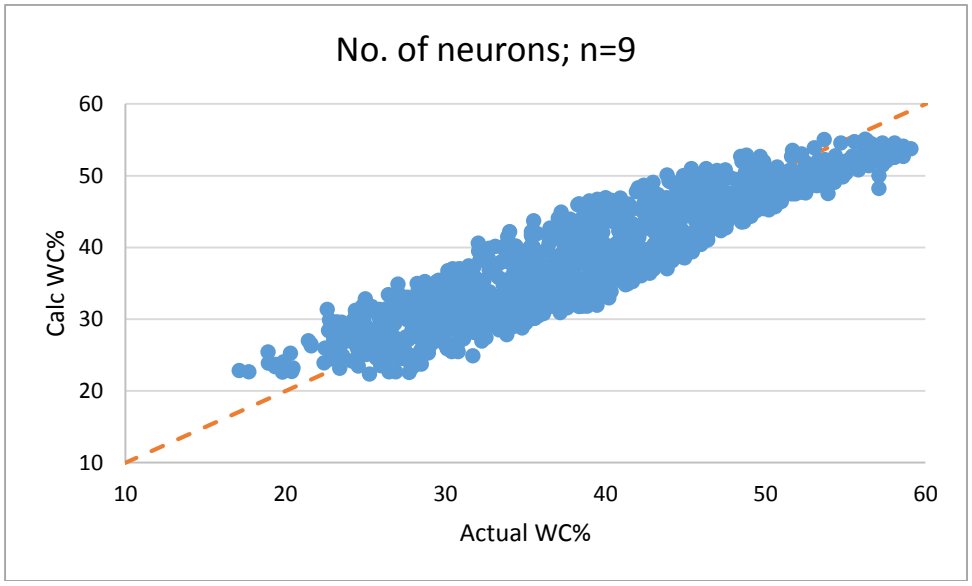


Figure B.23 Symmetric saturating linear (satlins) transfer function (n=9)

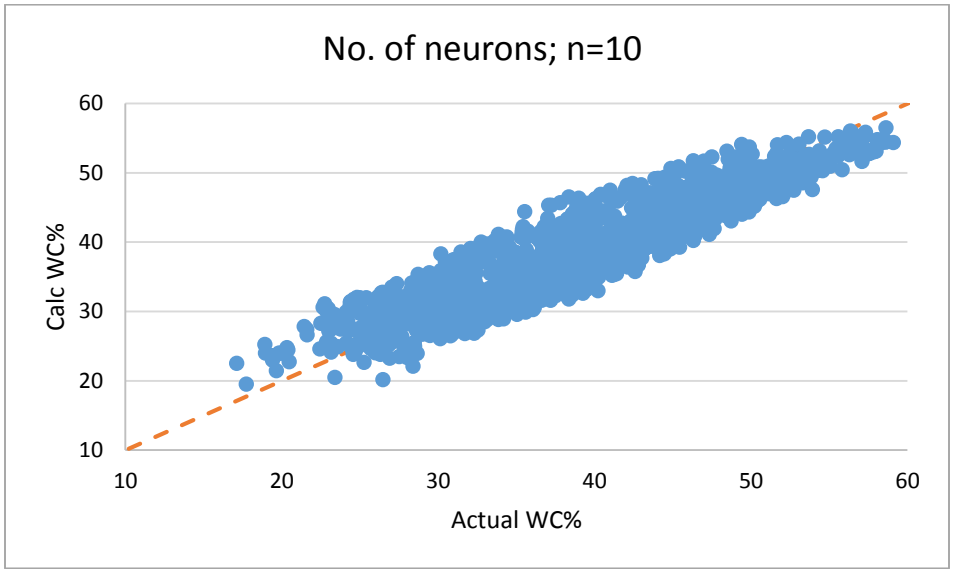


Figure B.24 Symmetric saturating linear (satlins) transfer function (n=10)

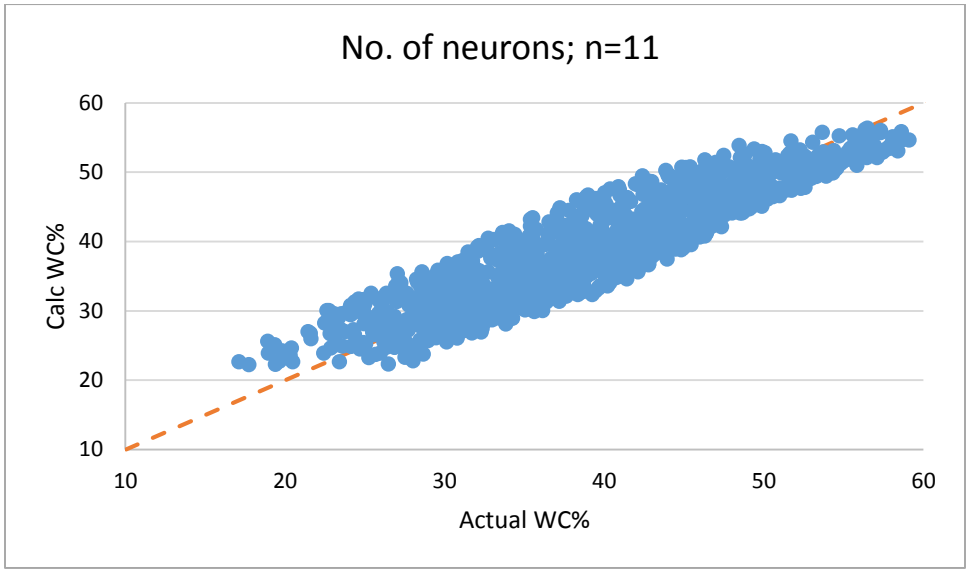


Figure B.25 Symmetric saturating linear (satlins) transfer function (n=11)

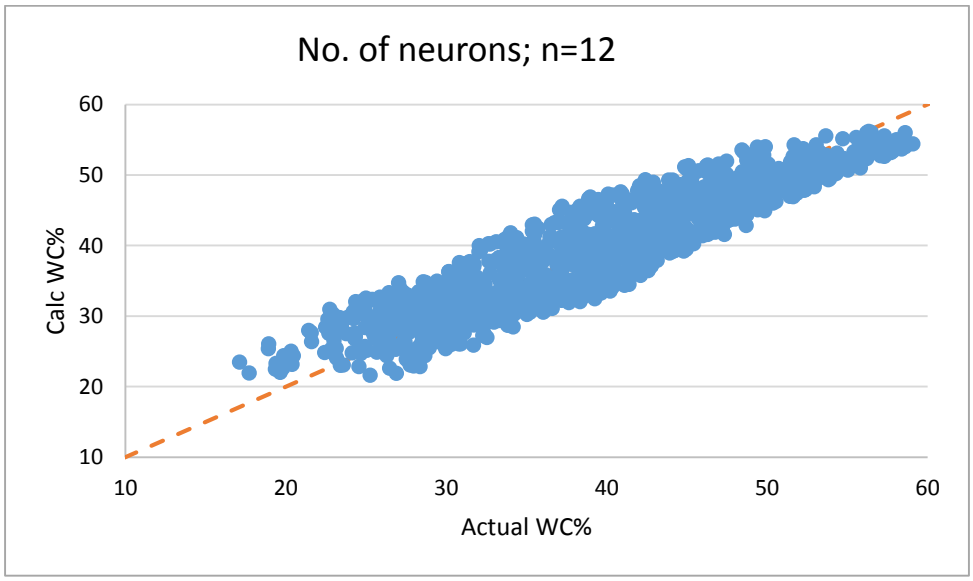


Figure B.26 Symmetric saturating linear (satlins) transfer function (n=12)

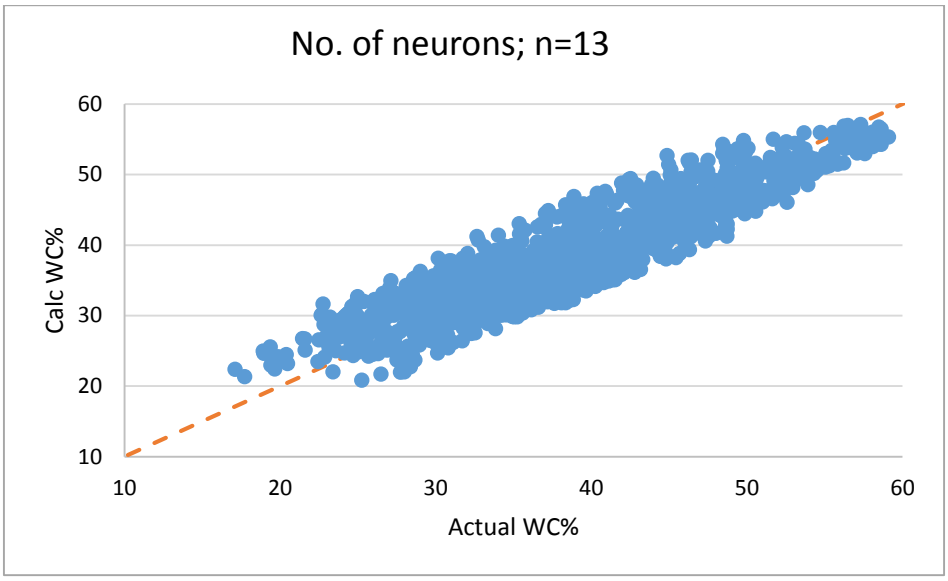


Figure B.27 Symmetric saturating linear (satlins) transfer function (n=13)

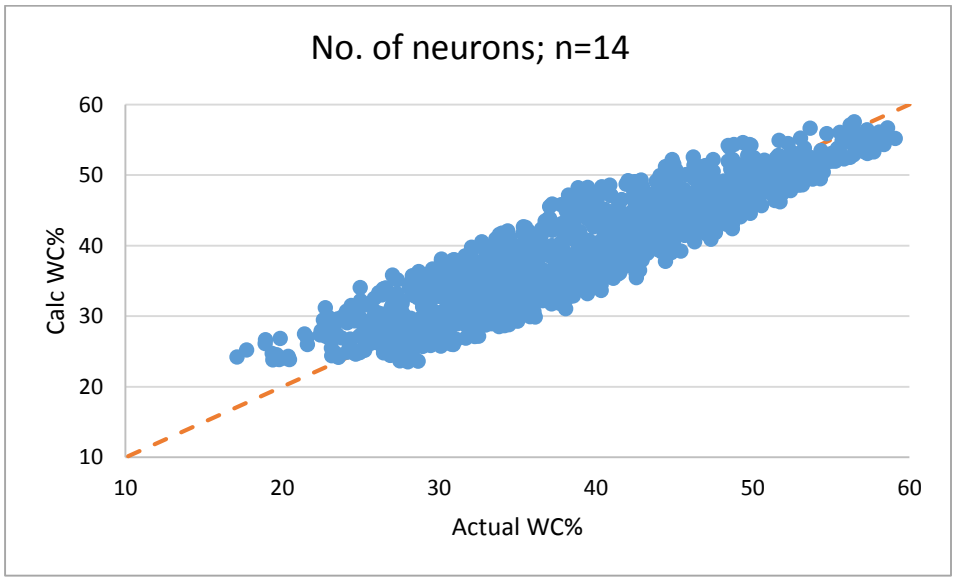


Figure B.28 Symmetric saturating linear (satlins) transfer function (n=14)

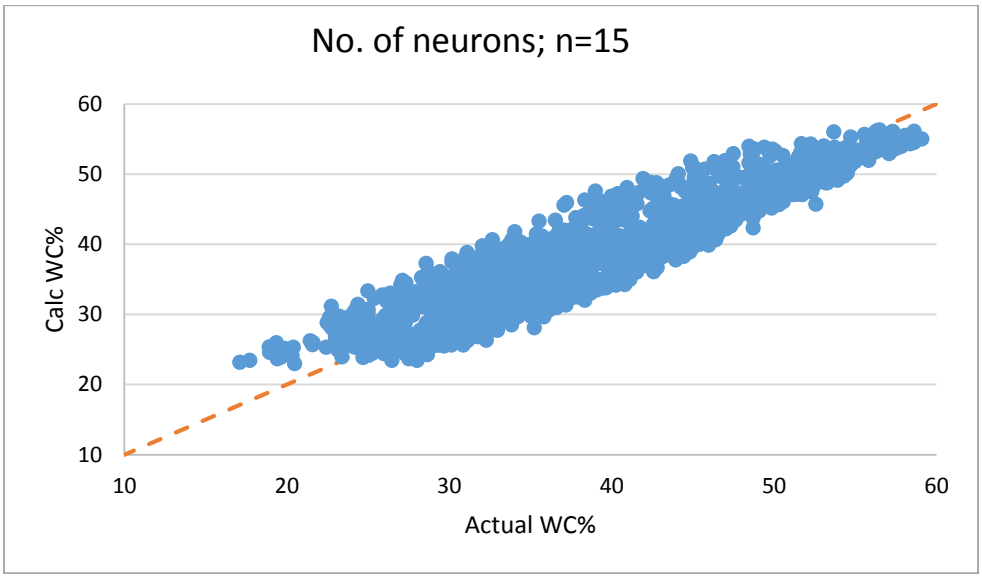


Figure B.29 Symmetric saturating linear (satlins) transfer function (n=15)

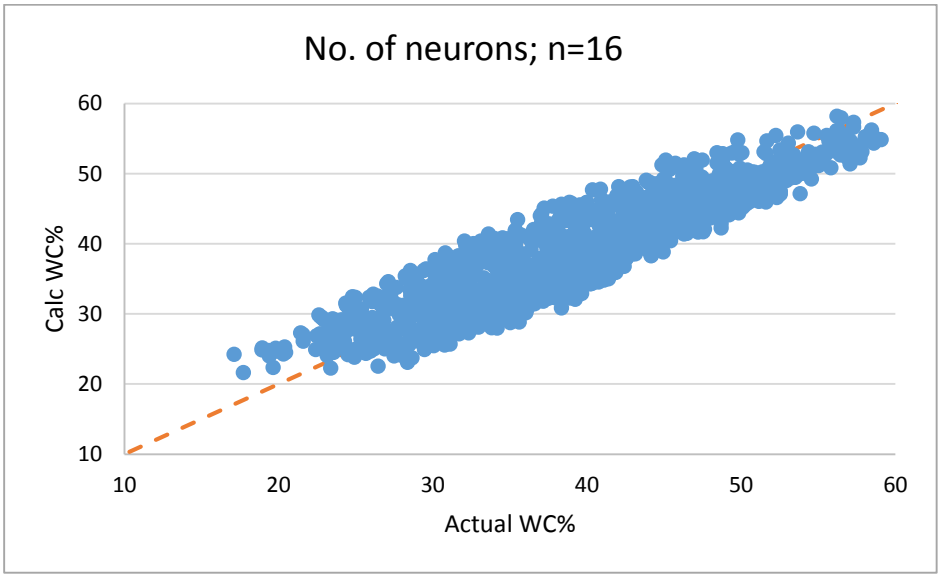


Figure B.30 Symmetric saturating linear (satlins) transfer function (n=16)

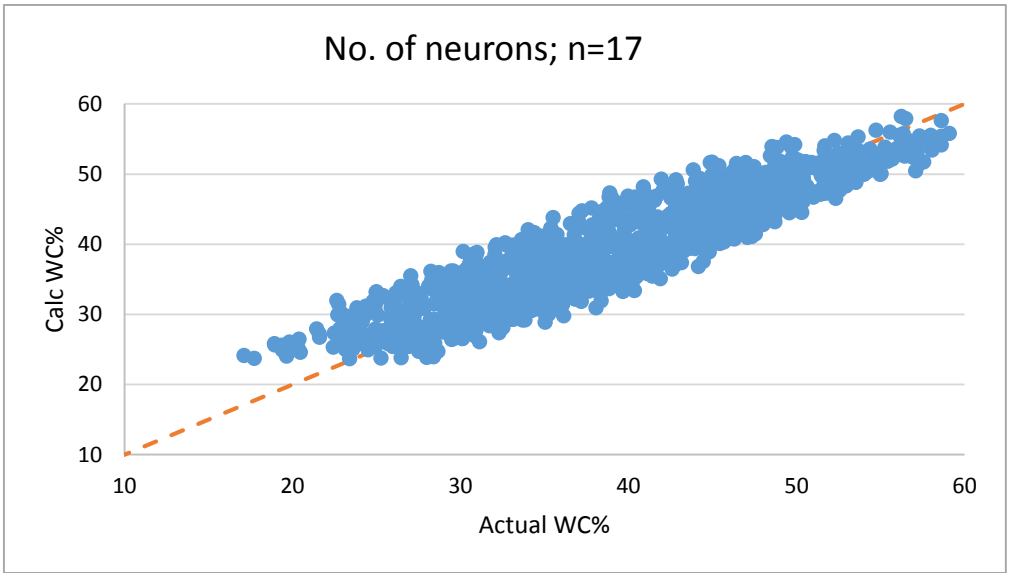


Figure B.31 Symmetric saturating linear (satlins) transfer function (n=17)

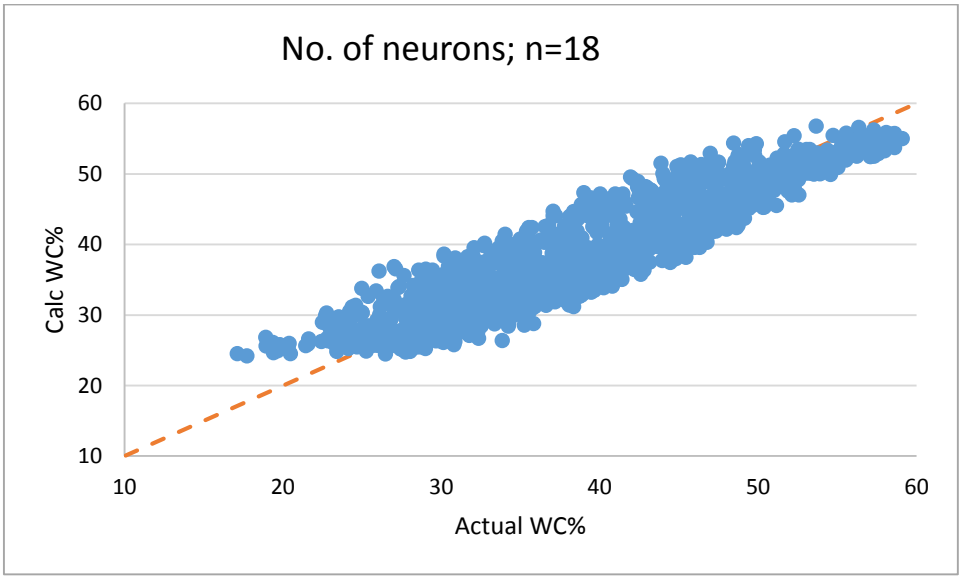


Figure B.32 Symmetric saturating linear (satlins) transfer function (n=18)

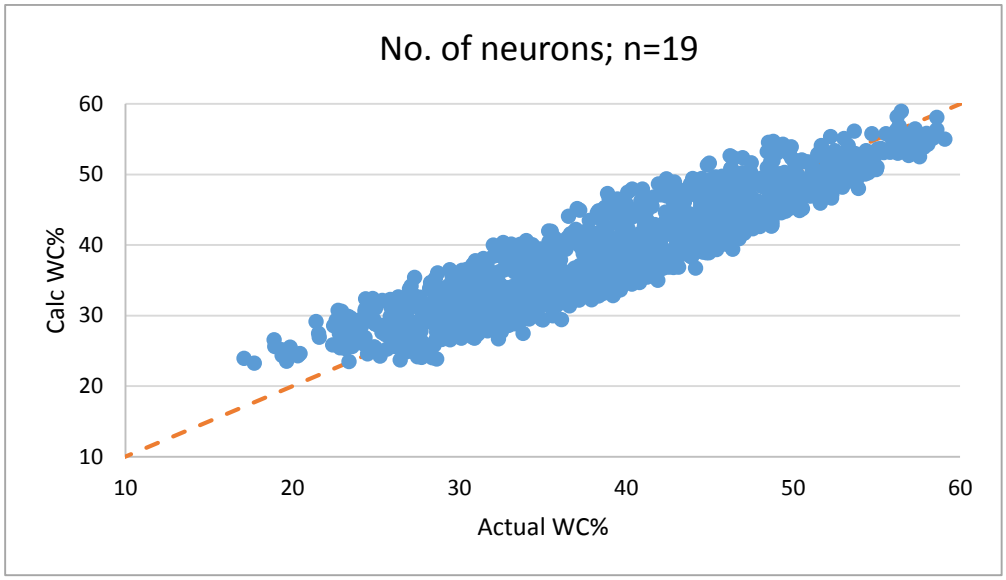


Figure B.33 Symmetric saturating linear (satlins) transfer function (n=19)

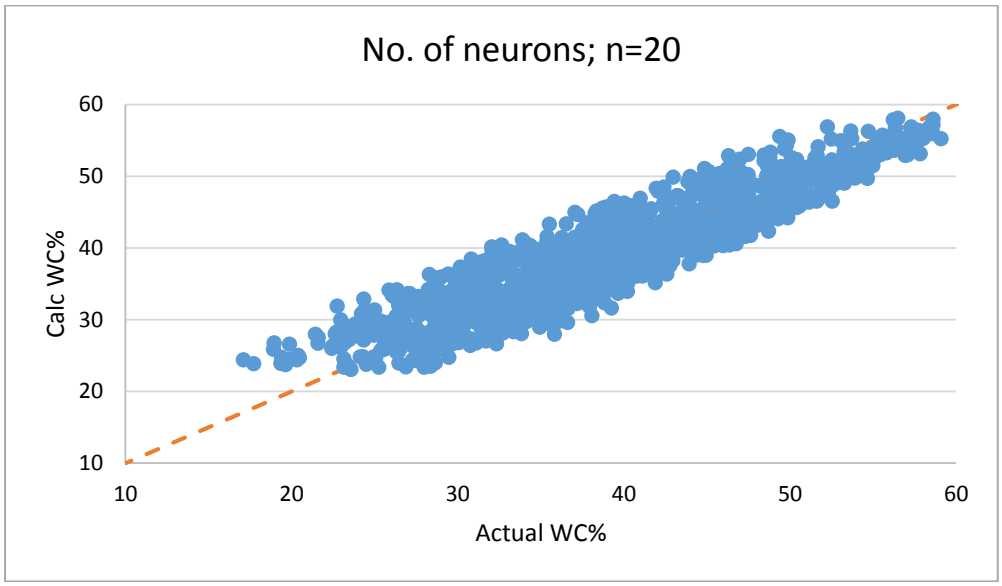


Figure B.34 Symmetric saturating linear (satlins) transfer function (n=20)

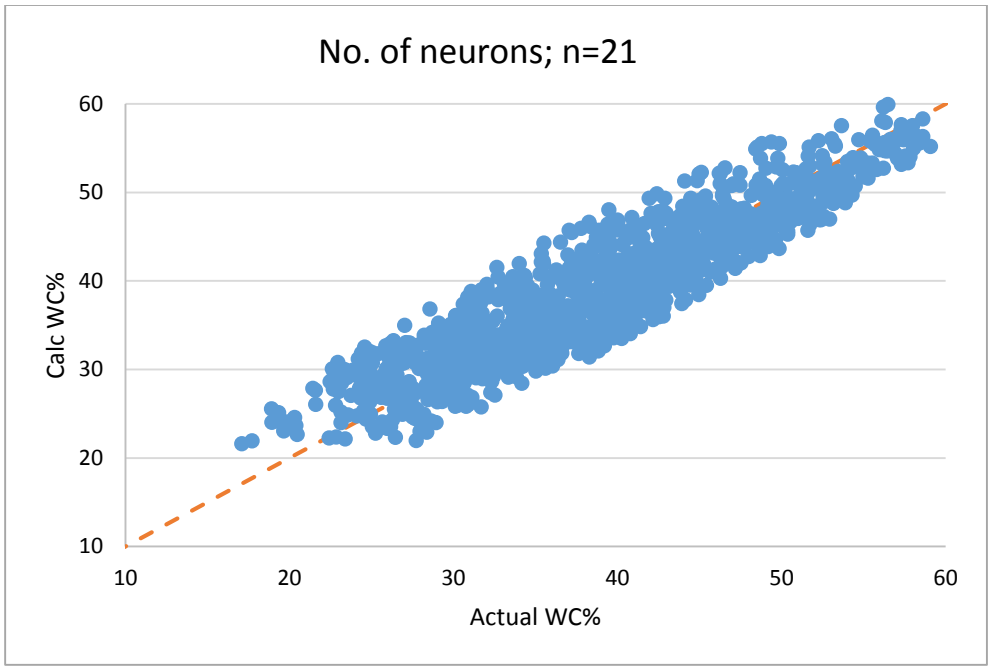


Figure B.35 Symmetric saturating linear (satlins) transfer function (n=21)

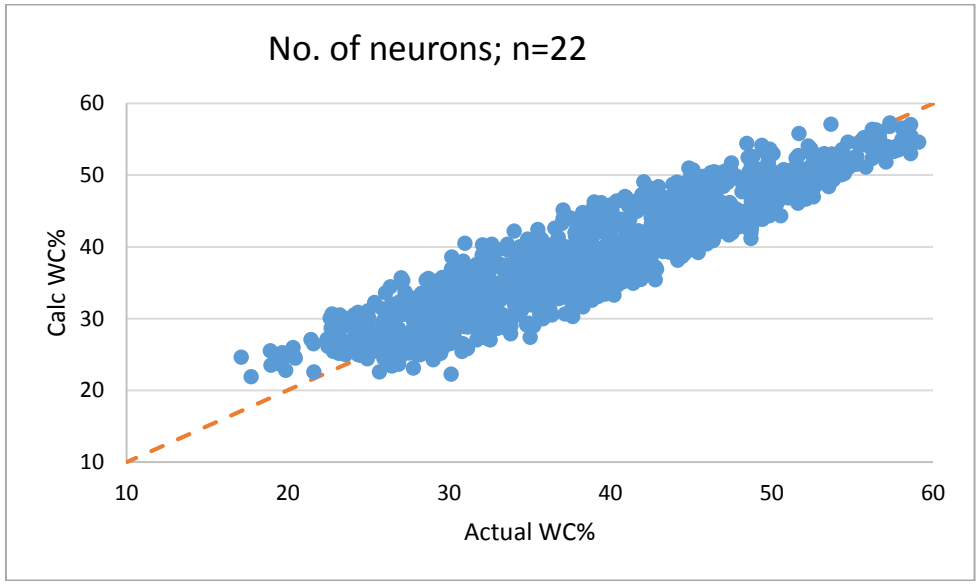


Figure B.36 Symmetric saturating linear (satlins) transfer function (n=22)

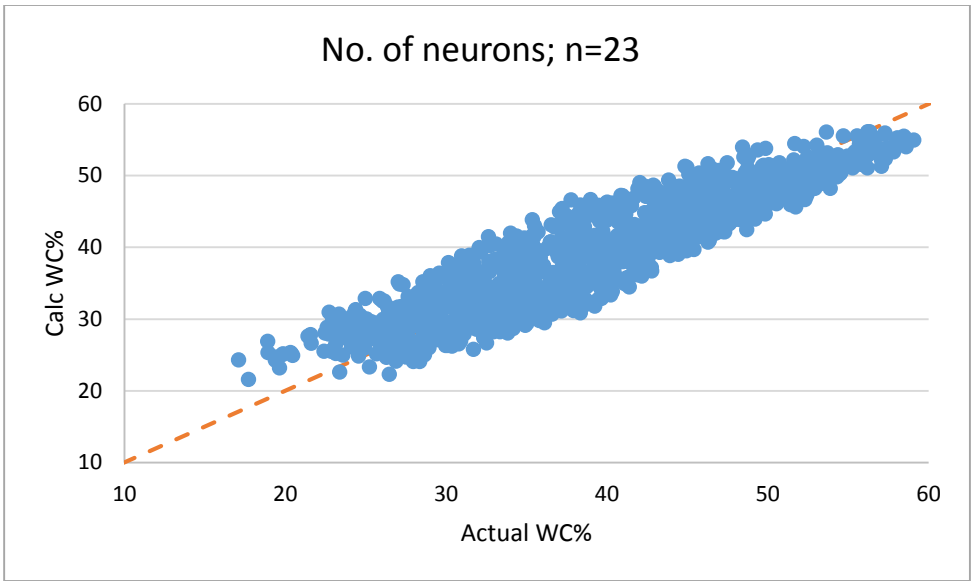


Figure B.37 Symmetric saturating linear (satlins) transfer function (n=23)

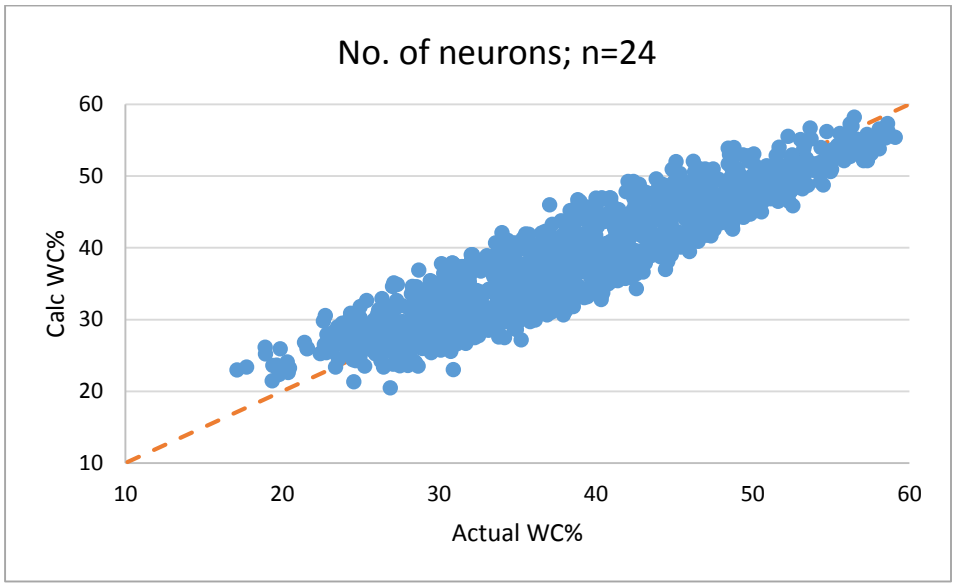


Figure B.38 Symmetric saturating linear (satlins) transfer function (n=24)

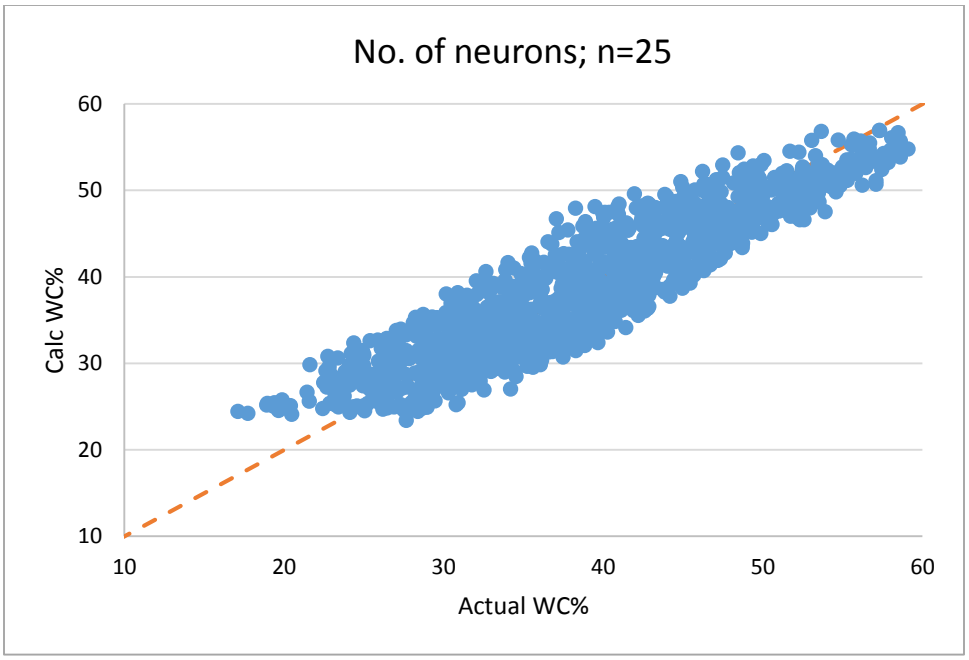


Figure B.39 Symmetric saturating linear (satlins) transfer function (n=25)

REFERENCES

Al-Amri, M, Al-Khelaiwi, F., & Al-Kadem, M. 2012. Advanced Utilization of Downhole Sensors for Water-cut and Flow Rate Estimation. Paper SPE 161063 presented at Abu Dhabi International Petroleum Conference and Exhibition, 11-14 November, Abu Dhabi, UAE.

Arsalan, M., Ahmad, T., Black, M., & Noui-Mehidi, M. 2015. Challenges of Permanent Downhole Water Cut Measurement in Multilateral Wells. Paper SPE 177665 presented at Abu Dhabi International Petroleum Exhibition and Conference, 9-12 November, Abu Dhabi, UAE.

Attia, M., Abdulraheem, A., & Mahmoud, M. 2015. Pressure Drop Due to Multiphase Flow Using Four Artificial Intelligence Methods. Paper SPE 175724 presented at SPE North Africa Technical Conference and Exhibition, 14-16 September, Cairo, Egypt.

Bansal, G. 2015. What is the difference between coefficient of determination, and coefficient of correlation?. Retrieved from http://blog.uwgb.edu/ on October 2018.

Breiman, L. & Cutler, A. 2018. Random Forests. Retrieved from https://www.stat.berkeley.edu on October 2018.

Camilleri, L., & Zhou, W. 2011. Obtaining Real-Time Flow Rate, Water Cut, and Reservoir Diagnostics from ESP Gauge Data. Paper SPE145542 presented at Offshore Europe, 6-8 September, Aberdeen, UK.

Cellos, H., & Wee, A. 1999. Multiphase-Flow Measurement System of High-GOR Applications. Paper SPE 54605 presented at SPE Western Regional Meeting, 26-27 May, Anchorage, Alaska, USA.

Dali, S. 2015. Role of Bias in Neural Networks. Retrieved from https://stackoverflow.com on October 2018.

Donges, N, 2018. The Random Forest Algorithm Retrieved from https://towardsdatascience.com/ on October 2018.

Economides, M. 1996. Production Engineering, Standard Handbook of Petroleum and Natural Gas Engineering. Gulf Professional Publishing, First Edition,1996

Ehtesham H., Putra, S., & Shammari, A. 2011. Coupled Facility and Reservoir Simulations to Optimize Strategies for a Mature Field. Paper SPE 147994 presented at SPE Reservoir Characterisation and Simulation Conference and Exhibition, 9-11 October, Abu Dhabi, UAE.

Al Enezi, S., Warlick, M., Almusabeh, M., & Kaba, A. 2012. Forecasting and Monitoring Water Cut Utilizing ESP Pump Discharge Pressures and Fluid PVT Analysis. Paper SPE 160886 presented at SPE Saudi Arabia Section Technical Symposium and Exhibition, 8-11 April, Al-Khobar, Saudi Arabia.

Falcone, G., Hewitt, G. F. and Alimonti, C., 2009. Multiphase Flow Metering: Principles and Applications Textbook, Elsevier Inc., Oxford, UK, First Edition 2009.

Fancher, G., & Brown, K. 1963. Prediction of Pressure Gradients for Multiphase Flow in Tubing. Paper SPE 440 published in Society of Petroleum Engineers Journal, Volume 3, Issue 01, Pages 59-69.

Farouk, H. 2013. Artificial Neural Networks. Retrieved from https://www.researchgate.net/ on October 2018.

Genolini, M., 2016. High Gas Multiphase Flow Meter (MPFM) ABB VIS. Retrieved from https://library.e.abb.com on March 2018.

Ghareeb, M., & Elgaghah, S. 2007. A New Correlation for Calculating Wellhead Production Considering Influences of Temperature, GOR, and Water-Cut for Artificially Lifted Wells. Paper IPTC 11101 presented at International Petroleum Technology Conference, 4-6 December, Dubai, U.A.E.

Guo, B., Lyons, W., & Ghalambor, A., 2007. Petroleum Production Engineering-A Computer-Assisted Approach. Elsevier Science & Technology Books, USA, First Edition, pages 57-62.

Hatton, G., Helms, D., Marrelli, J., & Durrett, M 1990. A New Microwave-Based Water-Cut Monitor Technology. Paper OTC 6426 presented at Offshore Technology Conference, 7-10 May, Houston, Texas, USA.

Holstein, 2007. SPE Reservoir Engineering and Petrophysics Volume V. From Petroleum Engineering Handbook, SPE Series, 2007.

Al-Kadem, M. 2014. A Decade of Experience with Multiphase Flow Meters. Paper IPTC 18162 presented at International Petroleum Technology Conference, 10-12 December, Kuala Lumpur, Malaysia.

Lake, L., 2003. Effect of Poor Data Quality on the Coefficient of Determination. Retrieved from https://www.researchgate.net/ on October 2018.

Li, K., Ren, X., Li, L., & Fan, X. 2011. A New Model for Predicting Water Cut in Oil Reservoirs. Paper SPE 143481 presented at SPE EUROPEC/EAGE Annual Conference and Exhibition, 23-26 May, Vienna, Austria.

Macro Trends, 2018. Crude Oil vs. Gasoline Prices - 10 Year Daily Chart. Retrieved from http://www.macrotrends.net on January 2018.

MathBits, 2018. Correlation Coefficient and Coefficient of Determination. Retrieved from https://mathbits.com on September 2018.

McCain, W. 1990. The Properties of Petroleum Fluids. PennWell Publishing Company, Tulsa, Oklahoma, Second Edition, Textbook pages 147-157.

Moshfeghian, M., 2014. Impact of Gas-Oil Ratio (GOR) on Crude Oil Pressure Drop in Gathering Systems. Retrieved from http://www.jmcampbell.com/ on March 2018.

Al-Mutairi, A., Khuzzan, S., Helal, R., & Raman, B. 2011. Enhancing Well Testing Performance by Installing a Reliable Water Cut Meter Along with Coriolis Flowmeter. Paper SPE 149117 presented at SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition, 15-18 May, Al-Khobar, Saudi Arabia.

Nasri, A., Al-Anizi, A., Al-Amri, M.., Al-Khelaiwi, F., & Al-Anazi, A. 2014. Multiphase Flow Meters Trial Testing in High GOR/GVF Environment. Paper IPTC 17422 presented at International Petroleum Technology Conference, 19-22 January, Doha, Qatar.

NFOGM, 2005. The Handbook of Multiphase Metering. Retrieved from http://nfogm.no/wpcontent on October 2018.

Nielsen, M. 2015. How the backpropagation algorithm works?. Retrieved from http://neuralnetworksanddeeplearning.com on October 2018.

Oglesby, K., Mehdizadeh, P., & Rodger, G. 2006. Portable Multiphase Production Tester for High-Water-Cut Wells. Paper SPE 103087 presented at SPE Annual Technical Conference and Exhibition, 24-27 September, San Antonio, Texas, USA.

Open University. 2018. Operation and Maintenance of Water Treatment and Supply Systems. Retrieved from http://www.open.edu on March 2018.

Peruzzi, T., & Krumanocker, E. 1999. High GOR Wells Optimization, Prudhoe Bay Practice. Paper SPE 54634 presented at SPE Western Regional Meeting, 26-27 May, Anchorage, Alaska, USA.

Pokharna, H. 2016. Introduction to Neural Networks we all need. Retrieved from https://medium.com/technologymadeeasy/ on October 2018.

Production Technology, 2017. Multiphase Flow Correlations. Retrieved from https://production-technology.org on March 2018.

Rao, B. 1998. Multiphase Flow Models Range of Applicability. Retrieved from https://pdfs.semanticscholar.org on March 2018.

Ray, S. 2015. 7 Types of Regression Techniques you should know!". Retrieved from https://www.analyticsvidhya.com/ on October 2018.

Safin, D. A., Korobkin, A., & Sitnikov, A. 2016. Horizontal Well Water Cut Estimation due to Water Coning in Heterogeneous Formations with Vertical Flow Barriers. Paper SPE

182047 presented at SPE Russian Petroleum Technology Conference and Exhibition, 24-26 October, Moscow, Russia.

Al-Saiyed, M., Warsi, S. A., Phillips, J., Gilani, S., Zareef, M., & Lievois, J. 2008. Measurement of Water Cut in Challenging Flow Conditions using Infrared Technology. Paper SPE 118038 presented at Abu Dhabi International Petroleum Exhibition and Conference, 3-6 November, Abu Dhabi, UAE.

Sharmal, B. and Venugopalan, K., 2014. Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images. From Journal of Computer Engineering, Volume 16, Issue 01, Version II (Jan. 2014), Pages 31-35.

Tseytlin, S., Tseytlin, D., Makarian, T., & Petrossov, V. 2016. New Technology of Optimization of Production of Liquid Hydrocarbons from Reservoirs Containing Oil or Condensate with High GOR and Oil Fringes of the Gas Formations (Russian). Paper SPE 181951 presented at SPE Russian Petroleum Technology Conference and Exhibition, 24-26 October, Moscow, Russia.

NOMENCLATURE

a, b, c, α, w: constants

A: Tubing cross-sectional area, in^2

b_1: Bias between input and the hidden layer of the neural network.

b_2: Bias between hidden and the output layer of the neural network.

D: Decline rate

E_a: Absolute Error, unitless

E_r: Absolute relative Error, fraction

E_{aa}: Average Absolute Error, unitless

E_{ar}: Average Absolute relative Percentage Error (AAPE), %

f_{oi}: Initial oil cut, fraction

f_w : Water cut, fraction

f_{wt}: Water cut as a function of time, fraction

GOR: Gas oil ratio, scf/STB

H: Well producing depth, ft

n: Number of neurons in the hidden layer, number.

N_p: Cumulative production, STB

P_{ds}: Downstream flowing wellhead pressure, psi

P_{wf}: flowing bottomhole pressure, psi

P_{wh}: Upstream flowing wellhead pressure, psi

Q; Gross liquid production rate, bbl/d

Q_L: Liquid flow rate, STB/day

Q_o: Oil flow rate, STB/day

Q_{oi}: Initial oil production rate, bbl/day

Q_t: Total production rate, bbl/day

R^2: Coefficient of Determination, unitless

r: Correlation Coefficient, unitless

T_{bh}: Bottomhole Temperature, °F

T_{th}: Tubing Head Temperature, °F

T_{wh}: Upstream flowing wellhead temperature, °F

T_{wf}; flowing bottomhole temperature, °F

w_1: Weights vector between input and the hidden layer of the neural network.

w_2: Weights vector between the hidden and the output layer of the neural network.

W_p: Cumulative water production, bbl

WC: Water cut, %

WOR: Water oil ratio, fraction

x: Input parameters

y: Output variable represents the actual value

$\hat{y}$: Output variable represents the estimated value

VITAE

Name	: **Mohammad Saeed Al-Kadem**
Nationality	: Saudi
Date of Birth	: October 13, 1988
Email	: kadem10rmd@hotmail.com
Address	: P.O.Box 4683, Ras Tanura, 31311 Eastern Province, Saudi Arabia
Education	: Bachelor of Science in Petroleum Engineering, 2011 King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
Professional Certifications	: Society of Petroleum Engineers (SPE) and Lean Six Sigma
Work Experience	: Reservoir and facility Engineering
Research Interests	: Intelligent Field technologies and Big Data utilization
Selected Publications	: Here are some papers published as main author:

- Real-Time Estimation of Well Drainage Parameters; SPE-163386 at KIPCE 2012 in Kuwait
- First Autonomous ICD Installation in Saudi Arabia - Modeling a Field Case; SPE Paper 177997 at ATS&E 2015 in Saudi Arabia
- Real-Time Estimation of Flow Rate in Dry Gas Wells – A New Field; SPE Paper 182754 at ATS&E 2017 in Saudi Arabia
- Integrating Peng Robinson EOS With Association Term for Better Minimum Miscibility Pressure Estimation; SPE Paper 192327 at ATS&E 2018 in Saudi Arabia